Language Variation and Contact
Phonetic and Phonological Aspects of Portuguese of Maputo City

By

Aurélio Zacarias Simango

SMNAUR001

THESIS

in

Sociolinguistics

Submitted to the Faculty of Humanities in Partial Fulfilment of the Requirements for the Degree of Master in Linguistics

UNIVERSITY OF CAPE TOWN

2010

SUPERVISOR

Professor Rajend Mesthrie

DEPARTMENT OF ENGLISH

LINGUISTICS SECTION

*O **R**ATO **R**OEU A **R**OLHA DA G**ARR**AFA DO **R**EI DA **R**ÚSSIA*

The mouse gnawed the cork of the bottle of the King of Russia.

PLAGIARISM DECLARATION

I know that plagiarism is wrong. Plagiarism is to use another's work and pretend that it is one's own.

I have used the **Chicago Convention** for citation and referencing. Each contribution to, and quotation in, this thesis from the work(s) of other people has been attributed, and has been cited and referenced.

This thesis is my own work.

I have not allowed, and will not allow, anyone to copy my work with the intention of passing it off as his or her own work.

Signature: _____

Date: _____

DEDICATION

In memory of my father **Mangambanine Zacarias Simango** who taught me that:

Behold how good and how pleasant it is for brethren to dwell together in unity!

It is like the precious oil upon the head, running down on the beard, the beard of Aaron, running down on the edge of his garments.

It is like the dew of Hermon, descending upon the mountains of Zion; for there the LORD commanded the blessing – life forevermore.

(Psalms, 133:1-3)

ACKNOWLEDGEMENTS

This dissertation is a realisation of my dream of being able to meet competently my people's linguistic needs and challenges. It is possible because of a combination of efforts made by many people and institutions to them I will be in debt all my life. I, therefore, would like to take this chance to express my gratitude to all of them especially to hose who funded and to those from whom I received my training. I would like, hence, to record my sincere gratitude to the generosity of the Lestrade Scholarship from the Postgraduate Centre & Funding Office for having sponsored this study, as well as the NRF (National Research Foundation) via small grants from Professor R. Mesthrie.

My greatest and special debt of gratitude is to Professor Rajend Mesthrie, my supervisor and Head of the Section of Linguistics at the Department of English of the University of Cape Town, for his encouragement to me to start my specialization in one of the African varieties of Portuguese, the Mozambican Portuguese, and for having led me along my whole research on the phonetic and phonological variation in Mozambican Portuguese, and for his financial and moral support.

Another person whose influence was important to this study is Dr. Perpétua Gonçalves, Full Professor of Linguistics of Portuguese at Universidade Eduardo Mondlane, in Maputo/Mozambique, for her useful suggestions and support on the main literature and research on Portuguese varieties, especially the Mozambican and the Brazilian Portuguese. I am grateful to Professor J. K. (Jack) Chambers, Department of Linguistics, University of Toronto, for having sent me his publication in language acquisition, whose general principles in dialect acquisition guided this research; and to Dr. Clarissa Sureck-Clark for her master's thesis in dialect acquisition in Brazilian Portuguese.

This research has been made possible largely because of existence of linguistic survey on Mozambican Portuguese, the so-called Overview of Oral Portuguese of Maputo (*Panorama do Português Oral de Maputo – PPOM*). I would like to express my deepest gratitude in this regard to Dr. Albertina Chachuaio of National Development Institute of Education in Maputo/Mozambique (*Instituto Nacional do Desenvolvimento da Educação - INDE*) for having allowed me to use this linguistic data and other useful related information. I am also

deeply grateful for many colleagues for their multifaceted support. My special thanks in this regard go to my friends Izidine Pinto, for his patience to explain some foundations on general statistics needed for my analysis; Isham from the Language Laboratory of the School of Languages & Literatures of the University of Cape Town, for the technical assistance to convert all data from the analogue to digital system; Nuroo Hoosain-Ismail and Dianne Steele, from the Knowledge Commons, for the facilities in order to work peacefully at the Knowledge Commons; and to all my classmate colleagues for our discussions during and out the classes. My special gratitude is to Robyn Stephenson, and Alida Jacobs, former and current research assistants of the Section of Linguistics, respectively, for their administrative assistance. Particular thanks are due to Miss Faiza Steffenson the Executive Administrative Assistant of the Department of English of the University of Cape Town for her patience and institutional support.

However, of all people, I owe the greatest debt of gratitude to my nephew Adelino for having rescued me to the second life and to my spouse and friend, Joaquina (Quinita) Carlos José Esteira Simango, without her love, and unlimited indulgency, especially when I became desperate because of my sickness, this work would not have been completed. She has not only encouraged me to face the challenges of my life, but also has managed to survive very far from my support and love for many years.

Thanks are also due to my Mommy Raquelina (Xiluvane) Pedro Muianga, for having given me the natural life; and to my siblings for their understanding.

For all of them I would like to express sincerely and deeply my gratitude in my language of confidence, Xichangana: *Khanimambo*, which means thanks.

ABSTRACT

The main goal of this study was to determine the extent to which (some of) Chambers' (1998) "Eight Rules of Dialect Acquisition", also discussed by Surek-Clark (1998) in her study of Brazilian Portuguese speakers, apply to Mozambique Portuguese learners and if sociolinguistic factors such as age, education, residence and sex, play a significant role in allophonic distribution and sociolinguistic variation in Portuguese in Mozambique, taking into account community-based patterns of use.

The data used in this study is part of Panorama of Oral Portuguese of Maputo "*PPOM - Panorama do Português Oral do Maputo*", a linguistic survey comprised of individual interviews and group interviews carried out in 1997 in region of the City of Maputo and its surroundings undertaken by Christopher Stroud and Perpétua Gonçalves (1997).

The focus of this study is on the European Portuguese Norn socio-phonetic variable /ʀ/ represented by [ʁ] as uvular fricative, and also found in some dialects as alveolar trill represented by [r], and the tap or alveolar flap [ɾ], both in complementary distribution. The focus is also on the lateral palatal [ʎ] (Cruz-Ferreira 1999:126, Mateus et al. 1983:529 and Martins 1988:78). The aim is to describe the patterns of variation between alternative forms in language use, seeking to test the hypothesis that socio-phonetic variation of /ʀ/ and /ʎ/ is associated with age, sex, education and place of residence.

The results of this study were obtained from a quantitative analysis using VARBRUL, (Cedergren, H.; and Sankoff, D. 1974; Tagliamonte, S. 2006), a computer-based statistical program. They show that the sociolinguistic patterns involving uvular fricative [ʁ] or alveolar trill [r], tap or alveolar flap [ɾ], and lateral palatal [ʎ], lateral alveolar [l], or lateral fricative [ɬ], in Mozambican Portuguese, are strongly affected by age. Residence and sex play a moderate role, while education has the weakest effect.

LIST OF ACRONYMS

AAVE	African American Vernacular English
DU (1-7)	Distrito Urbano codified by decimal numbers, from 1 to 7 indicating the number of the districts of Maputo
EP	European Portuguese
INDE	Instituto Nacional do Desenvolvimento da Educação
INE	Instituto Nacional de Estatística
IPA	International Phonetic Alphabet
L1	First Language
L2	Second Language
MP	Mozambican Portuguese
NELIMO	Núcleo de Estudo de Línguas Moçambicanas
PPOM	Panorama do Português Oral do Maputo
SADC	Southern African Development Community
SIL	Summer Institute of Linguistics
SNE	Sistema Nacional de Educação

LIST OF ABBREVIATIONS

Bal	ciBalke
Cha	xiChangana
Chu	eChuwabo
Cop	ciCopi
GoldVarb	Multivariate Analysis Application Computer Program ran by (Rand & Sankoff 1990)
Ma	eMakhuwa
Mak	shiMakonde
Man	ciManyika
Mwa	kiMwani
Nda	ciNdau
Nya	ciNyanja
Nyu	ciNyungwe
Rho	xiRhonga
Sen	ciSena
Ton	giTonga
Tsh	ciTshwa
Ute	ciUtee
VarbRul	Variable Rule
Yao	ciYao

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF MAPS

CHAPTER 1
INTRODUCTION

This chapter consists of an introduction to the thesis and a brief formulation of the problem and the research aim, plus a brief overview of language contact and variation as fields of sociolinguistics. It ends with a statement of the hypothesis.

1.1 Socio-historical background

The colonial history of Mozambique is an eloquent example of language expansion and contact in Africa, as a consequence of the Portuguese discoveries. One of the most evident consequences resulting from this contact and reciprocal influences is undoubtedly language variation in Portuguese and the autochthonous languages. In fact, besides commerce, considered as its essential motivation, the Portuguese discoveries will also have a great impact on cultural and linguistic relations with the new lands, cultures and languages. It is also shown that, in order to solve linguistic constrains emerging from the contact with new cultures and languages in Africa, the Portuguese used local indigenes as interpreters. In this regard, Cardeira (2006:66) argues, "these contacts constitute the initial chapter of a history that the Portuguese will write in the following centuries, diffusing the Portuguese language in the new discovered lands". This is also the phase in which Portuguese was transported from Portugal to Africa. However, as will be discussed later in this section, contrary to Cardeira's (2006:65-66) conclusion, in any case this contact will lead to the replacement of the autochthonous languages by Portuguese.

This section aims to give a socio-historical background of the Portuguese in Mozambique in order to understand the development of a language change in the context of social life of the community in which it occurs. For instance, although the first occurrence of Portuguese in Mozambique can be associated with the Portuguese expansion in Africa, around 1414, it is only very late, during the 1930s where the majority of Africans were enrolled to the rudimentary education exclusively designated to them (Mendonça 1988:9). As she illustrated, a local newspaper 'O *Africano*', in one of its appearance noticed "…the subjects of His Majesty the king of Portugal don't speak Portuguese". Regarding to this,

among other authors, Mateus *et al.* (1983:24) associate the weak territorial expansion and restricted use of the language with the variation and change of Portuguese during the colonization period. Consequently, according to them, the Portuguese spoken in Angola and Mozambique has remained very close to that of Portugal because of the fact that only a very limited number of people used it during the 500 years of colonization (See Mateus et al. 1983:16).

Although I agree with these authors in some aspects, I consider that this apparent immutability of the Portuguese language cannot be taken for grant. First, during at least two centuries of initial colonization, Portugal did not have focussed linguistic 'norms' in the restricted sense of the word. It is only in the second half of XVI century and during the XVII century, with the poetic contribution of Sá de Miranda, where the interest in a more elaborated language showing codification and establishment of a Portuguese norm became consolidated (cf. Cardeira, 2006:71). Second, besides the lack of a norm, it is important to keep in mind that not all Portuguese settlers were always speakers of the same variety of Portuguese giving the diversity of their origins, on one hand, and on the other some of them were speakers of a very low level of proficiency. For example, there are cases in the countryside of Mozambique where the situation can be analyzed as a continuum. At one extreme, there are white parents who are gradually losing their command of Portuguese, while on the other end of the continuum the new generation ignore totally the Portuguese language for the benefit of African languages.

Exploring these aspects, the exploration of these two factors, namely the lack of an elaborated norm associated with a variety of degree of proficiencies of the colonial settlers may bring to light some further background details in order to deeper our understanding about the real conditions in which the Mozambican Portuguese started its variation.

Additionally, another important aspect to keep in mind raised by Gonçalves (1996:16) is the fact that until the 2nd half of the XVIII century the administration of Mozambique was made indirectly through India and the occupation of the territory was effective from 1918. This is also the period where the colonial system is relatively more concerned with the need of some kind of acculturation of the indigenous and the subsequent diffusion of the Portuguese language in a more formal and controlled environment.

The highest expression of this linguistic and cultural confrontation resulting from the violent contact with the colonization is the assimilationist politics, that is, the rupture with the cultural and linguistic universe legacy by the Africans and the adoption of the other imposed as an alternative for the prestige and social ascension (Mendonça 1988:12). However, as the result of the assimilations' politics, as a process of dialect acquisition, and by extension, the Portuguese of the '*assimilado*' is not a European variety but it is also not a totally indigenous Mozambican Portuguese either. Probably this is one of the starting points of the whole process of variation and change of the Mozambican Portuguese. In fact, on the one hand, the '*assimilado*' will never break with his/her origins to which the only access remains the African language. On the other hand, the Portuguese language acquired as a foreign language and mainly in an artificial school environment is to the African minority, literate and assimilated, the language of 'civilization and modernity' and for social ascension. Therefore, its uses are restricted to the professional needs of communication with Europeans. (Mendonça 1988:12-13).

According to Gonçalves (1996:27), based on the results of the first General Census of Population held in 1980, five years after the proclamation of Mozambique as an independent state, Portuguese was spoken by approximately 25% of an estimated population of 12 million, of which almost 2% spoke it as L1.

After independence and continuing a practice initiated during the process of liberation of Mozambique, the Portuguese, previously declared the language of unity, is cumulatively the official language and the language of wider communication. Alternatively, in some political events local languages are used for interpretation. As in the colonial period, the radio is the only space in which the use of indigenous languages is partial and officially allowed. The use of Bantu languages/mother tongues of the majority is totally prohibited at the school (Gonçalves 1996:17). This is in fact the phase of the real spread of Portuguese and its massive use that has irreversibly determined the variation characteristic of the language. At the sociolinguistic point of view, this historical moment is symbolically proclaimed through the following notably words by Samora Machel (1981, the first president of Mozambique between 1975 and 1986, quoted by Gonçalves 1996:19) as the period of the enrichment of the languages "*Enriqucer a Língua*".

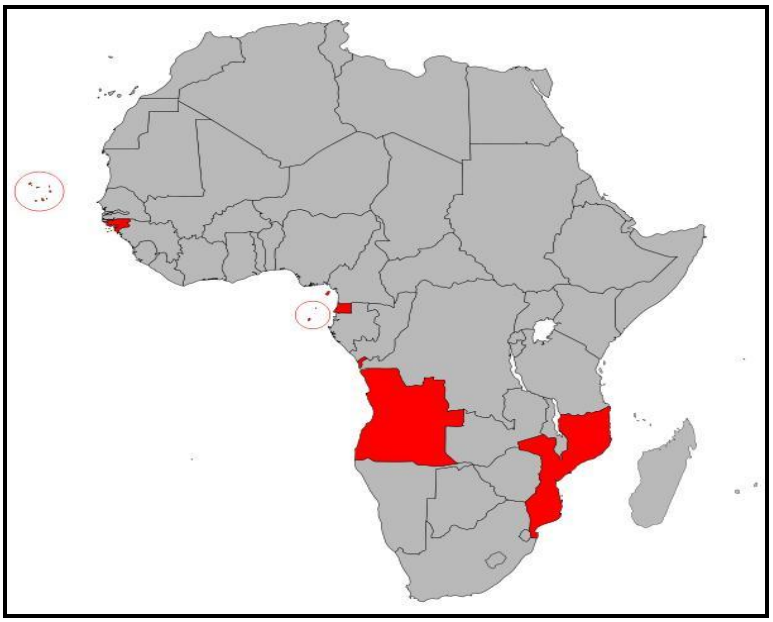
Consequently, the result of this process of language enrichment is a rapid growth of the number of Mozambican speakers of Portuguese in a situation of a quasi totally ignorance of the European norm. In my point of view this is the starting period of what I call massive 'dialectalization' of Portuguese in Mozambique in analogy to the notions of '*indigenization*' or '*nativization*' (cf. Firmino 2002:211 quoting Moag 1982 and Kachru 1982 and Mesthrie 1992:3, citing Kachru 1983a:2-3), or '*Africanization*' (Bokamba 1983). Contrary to the period of assimilation in which indigenous speakers had to imitate the European norm, in the process of 'dialectalization' the Portuguese language acquires new words, accents, and even different meanings. Therefore, according to Mesthrire (1992:1), given the fact that the dominant world languages acquire a multiplicity of forms in different host territories, sociologically it is more appropriate to refer to the resulting situations in terms of "language families" or "language complexes", in order to cover the emergent varieties with different histories functions and structural characteristics.

The following factors are relevant to the spread of Portguese and consequently the emergence of new varieities, namely: (a) the intense use of Portuguese in the mass media, especially on radio, and more recently television and in the information technology systems (ITS); (b) the extraordinary democratization of the Mozambican school subsequently to the period of the independency proclamation in 1975; and (c) the growth of in-migration in the cities, stimulated by the civil war. All of these aspects have had an impact on interpenetration of the process of dialect acquisition in Mozambican Portuguese.

The immediate consequence of these social alterations of the Mozambican society is obviously the variation in the spoken language, in the case, the spoken Portuguese. Nevertheless, similarly to the Brazilian case described by Castro (2006:29, quoting Teyssier 1982:79), what seems to be important to me in for future Mozambican sociolinguistics is to study the importance of dialectal variation compared to the more normative expectations of formal society. In many cases it is becoming difficult to pin down a specific construction, vocabulary item, or any phonetic feature as belonging exclusively to a geographical or socially limited group.

1.2 Some sociolinguistic aspects of situation of Mozambique

Portuguese, a Romance language with its origins in Europe, today is spread in a wider range of geographical discontinuous spaces inhabited by multiethnic and multilingual groups around the world (in this study I do not draw a distinction between the terms 'multilingual' and 'bilingual'). In 2005 Portuguese was spoken as mother tongue by nearly 200 million people, of which only 10 million live in Portugal, 185 million in Brazil and a significant number in the capital cities and their surroundings in Angola and Mozambique (Castro 2006: 11). The following map shows the five Portuguese-speaking African countries, also referred to in Portuguese by the acronym *PALOP* (*Países Africanos de Língua Oficial Portuguesa* 'African Countries of Portuguese Official'), where Portuguese language is the official language, namely: Angola, Cape Verde, Guinea Bissau, Mozambique and São Tomé and Príncipe.



Map 1: Portuguese-speaking African countries highlighted
(Source: http://en.wikipedia.org/wik)

Despite the persistent adoption of the European norm, recent studies show evidence of the formation and development of local varieties of Mozambican Portuguese (henceforth, MP), also called Portuguese of Mozambique (Gonçalves 1996). However, the varieties of Portuguese in Mozambique do not yet have a well-determined and established standard

variety. Instead, it comprises a multiplicity of distinct regional varieties (Hyltenstam & Stroud 1993, cited by Stroud & Gonçalves 1997:2). MP thus has much in common with Brazilian Portuguese.

According to Gonçalves (2003:50-51) the changes in the Mozambican Portuguese variety result from natural processes of acquisition of a non-mother tongue in a context where native speakers of African languages do not have significant access to the target language, that is, the European norm. In fact, the increasing of education facilities dominated by Portuguese as the exclusive medium of instruction at the time of the independence of Mozambique was accompanied by a growth in number of speakers and the range of varieties and functions. This is a direct result both of an increase in the number of the students enrolled and the length of time of their exposure to the language, from the primary, through secondary to the tertiary level.

However, despite the major role currently played by schools in the process of expansion of Portuguese in Mozambique (Stroud and Gonçalves 1997), it is undoubtedly not only in the classroom that people learn a language. They acquire it from different sources to communicate their horizons and beliefs, and as noted by Firmino, G. (2002:218) and Platt et.al (1984:4), schools are not islands in their respective communities. On the other hand, despite the use of standard European Portuguese (henceforth, EP, Gonçalves 1985) textbooks, most of the teachers speak already a distant variety from that of EP.

Furthermore, the language variation of MP is also one of the most evident consequences resulting from the contact between Portuguese and the autochthonous Mozambican languages. Consequently, as in other African and Asian contexts, many children who are native speakers of diverse African languages acquire a type of Portuguese that is already undergoing a process of variation and change. This situation influences negatively on the key question of language learnability at schools, since it becomes difficulty to define what to learn and how to learn.

One of the most salient differences between speakers of MP is at the level of pronunciation. In effect, speakers of MP variant from different regions of the country are

easily recognized as of Maputo, Nampula, Zambézia, Gaza or Cabo Delgado origin. Map 2 below indicates the 10 provinces of Mozambique, plus the capital, the city of Maputo.

Map 2: Provinces of Mozambique, plus the capital Maputo

Source: http://www.lib.utexas.edu/maps/africa/mozambique_pol95.jpg

However, with the exception of the morphsyntactic level, up to now, little work has been done on phonetic variation of MP under a close variationist perspective across different ethnolinguistic groups. Such studies are not sufficiently systematic or detailed to allow solid conclusions (see Mateus et al. 1983; Gonçalves 1996; Stroud and Gonçalves 1997).

This study discusses how the Eight Principles of Dialect Acquisition postulated by Chambers (1998) apply to MP learners. These principles determine the acquisition of dialect features by immigrants from one dialect region to another. The study also investigates

whether sociolinguistic factors such as age, education and mother tongue play a significant role in variant allophonic distribution and other ongoing process of variation of Portuguese in Mozambique.

In order to analyse this, I discuss oral production samples of bilingual Portuguese speakers in Maputo run by VARBRUL. More specifically, this study explores:

(i) To what extent do the Eight Principles of Dialect Acquisition postulated by Chambers (1998) account to MP learners?
(ii) Do sociolinguistic factors such as age, education and mother tongue play a significant role in variant allophonic distribution and other ongoing process of variation of Portuguese in Mozambique?

A new relevant dimension in the discussion is brought in this study by two key sociolinguistic variables not directly addressed neither by Chambers (1998) nor by Surek-Clark (1998) in their separate studies: the individual education background and the mother tongue, which is completely different from the target language. The study takes into account the community-based patterns of variation in relation to the interaction among different social factors.

In his study, Chambers' (1998) objective is to account for which features from the source dialect, i.e., L1 are lost and which features from the target dialect migrating individuals acquire. He considers the age at the 'destination dialect' (or the target dialect) is spoken. Inspired by these principles Surek-Clark (1988) investigates their implications for speakers of Brazilian Portuguese (henceforth, BP). Besides Chambers' (1998) variables of age and time, she also included two new variables, namely dialect prestige and speakers' ethnicity.

Traditionally, mainstream linguistic theory has given privilege to native speakers as its main source of language data, neglecting the non-mother tongue users (e.g. Chomsky 1965). On the contrary, Ferguson (1983) argues that at the level of universal explanatory principles or at the level of general theory of language, both native speaker and non-native kind of language use is important to the linguistic research. This follows from the fact that language variation between two languages is as important as it is in one single language to

account “phenomena like language acquisition, language convergence or language shift in relation to the learnability of natural languages and the nature of language change” (Ferguson 1983: viii).

In his analysis of the results, Chambers is based on the distinction between two key concepts, namely ‘accommodation’ and ‘acquisition’. Chambers defines accommodation as a transitory linguistic behaviour comprising “individual’s modifications of accent and dialect as a direct response to a particular interlocutor in a particular setting” (Chambers 1998:147).

In his turn, Trudgill (1986:40, quoted by Chambers 1998:147) distinguishes between ‘long term’ and ‘short term’ accommodation. He defines long term accommodation as a basic level of dialect adjustment ‘maintained’ by the individual in all transactions in the contact area, while short term accommodation is characterized as a kind of a ‘transitory’ adjustment above the latter made by the speaker in response to a particular social circumstance. Therefore, according to Trudgill, acquisition is a form of accommodation wherein the speaker adopts an alien linguistic variety so frequently that it becomes a permanent part of his/her accent or dialect by replacing original features.

Taking this Trudgillian perspective, similarly to Chambers’(1998) subjects’ native Canadian English, the subjects of this study, largely native speakers of Bantu languages, are engaged in a sort of long-term accommodation leading to a process of language acquisition of Portuguese.

The fact that the vast majority of people who use Portuguese in Mozambique speak it as a second language and are not directly exposed to the standard form of the Portuguese target language, results in a significant degree of variation at every linguistic level. This variation is, by and large, not controlled either by the education system or by the radio. To some degree, at least in the major urban areas, this also applies to the television and news press, where most of the teachers and the professionals in the communication enterprises don’t master the standard variety. Some socio-phoneticists call this level of language knowledge as ‘knowing what’ in contrast to “knowing how’ with reference to the knowledge natives are expected to have in order to control the use of their language and be able to communicate effectively with other speakers of your dialect, except in cases of pathological

impairment (Guitart 2005:16). The lack of one or both kinds of knowledge can be influenced by many social factors like age, education, linguistic background, or gender. For example, Guitart argues that bilingual speakers normally know only one of their dialects really well, although full bilingualism is not excluded.

Therefore, it becomes crucial to incorporate two new important variables in this research, not considered by Chambers (1998) and Surek-Clark (1998), namely:

(i) the bilingual background of the speakers, and
(ii) the acquisition of a second language rather than a second dialect.

In fact, these two aspects are related to one of the key language acquisition questions posed by Archibald (1995, xiii) "What do we know when we know phonology"? In other words, and related to this study, what a bilingual speaker learner is expected to know and reproduce about the phonology of a second language? This question cannot be divorced from considerations of the approach to language acquisition also referred to as language learnability, namely: (i) an account of the phonological knowledge of a native speaker and (ii) an account of the acquisition of that knowledge. This question will be discussed in chapter 3, dedicated to literature review.

Like the work of Surek-Clark (1998), this study will also address the question of dialect prestige of the variety being studied. Indeed, according to Stroud and Gonçalves (1997), Portuguese in Mozambique is expanding in terms of geographical spread and prestige. They argue that one of the most salient futures in the speakers' linguistic behaviour is what they call "linguistic uncertainty" (Stroud and Gonçalves 1997: 3, quoting Albarran 1991: 87). According to the authors, such linguistic uncertainty results, on the one hand, from the status given to Portuguese, which requires speakers to reorganise and adjust their grammar in order to converge to the EP norm. On the other hand, they consider that this linguistic uncertainty is related to the speakers' mother tongue competence, which is not socially valued.

An important aspect not discussed in this study is whether phonetic variation taking place in MP is a result of mutual phonological influences across ethnic boundaries, that is, if

speakers of different Bantu languages influence each other in the way they acquire Portuguese. In fact, in order to account for phonological influences across ethnic boundaries it is crucial firstly to delimit the phonological ethnic boundaries individually. Wolfram (1974) refers to evidence of dialect influences across ethnic boundaries. He points to the case of African-American Vernacular English (henceforth AAVE) regarding the ways in which contact with speakers of AAVE influenced the speech of Puerto Ricans in New York (cited by Fought 1999:6).

1.3 The data of this study: Social groups studied

A crucial question raised by Fought concerns the sufficiency of social factors traditionally used in studies of sound change, such as age, to account for sociolinguistic variation in a given community (Fought 1999:7). Therefore, alternative approaches have been based on an increasing use of ethnographic techniques in sociolinguistic inquiries. In this regard, Eckert (1991:213), quoted by Fought (1999:7) argues that:

> The use of ethnography in the study of variation allows the researcher to discover the social groups, categories and divisions particular to the community in question and to explore their relation to linguistic form.

The data used in this study is part of PPOM, a linguistic survey comprised of individual interviews, as elicited data, and informal group discussions, as more spontaneous tasks, carried out in 1997 in region of the City of Maputo and its surroundings. According to Stroud and Gonçalves (1997: 4) the PPOM Project aims primarily to build a comprehensive corpus of Portuguese speakers' in the city of Maputo, which is considered the major centre of Portuguese in the country. Therefore, and contrary to traditional practices of sample construction in well-established monolingual societies, the Maputo linguistic survey was informed by the attempt to show how the varieties of Portuguese of Mozambique, both as L1 but chiefly as L2, are socially structured, taking into account a range of social, linguistic-demographic and cultural factors.

The study uses only sociolinguistic structured personal interviews drawn from the PPOM, which were deemed particularly relevant and significant for sociolinguistic study. Therefore,

data collected by the use of the technique of free meetings where the informants choose their partners to talk about different issues were excluded from this study. In Chapter 4, on the methodology, I will return to this question.

Although there is a lack of variationist studies specifically related to the phonetic and phonological levels, my preliminary empirical observations allowed me to identify certain regional or national phonetic features of Mozambican Portuguese.

The set of symbols used in this study reflects the International Phonetic Alphabet principles, commonly abbreviated as 'IPA'. However, despite this fact, in order to describe the Portuguese sounds unambiguously and to make them comprehensible enough, many IPA symbols were drawn from different sources especially from Portuguese scholars, and they show also the dynamic of phonetics as a science and the its history. For instance, the European Portuguese form /ʀ/ described in this study as uvular fricative and represented by [ʁ], is also found as an alveolar trill in some dialects and represented by [r] (Cruz-Ferreira 1999:126 and Ladefoged et al. 1996:426). It is designated as 'vibrant velar' or vibrant multiple and represented by [ʀ] and by [ř] (Mateus et al. 1983:529). The tap or alveolar flap [ɾ] (Cruz-Ferreira 1999:126 and Ladefoged et al. 1996:426) it is referred to as vibrant alveolar simples and represented by [r] (Martins 1988:78)[1].

The observations are as follows:

The lack in all positions of the following voiced stops (also called "occlusive" or "plosive") as shown in Table 1.

Written forms	European Portuguese	Mozambican Portuguese	English
bar	[b]ar	[b]ar; [p]ar	'bar'
galo	[g]alo	[g]alo; [k]alo	'cock'

Table 1: Voiced stops absent in some MP varieties

[1] The nomenclature and phonetic symbols used in this study also includes (1) Moulton, William G. (1962) and (2) Sitoe, B. e Ngunga A. (eds.), (2000).

Note that in standard Portuguese, these phonemes are contrastive, that is, they form minimal pairs as shown in the following examples in Table 2.

Written forms	European Portuguese	English
par / bar	[p]ar / [b]ar	‘par’ ; ‘bar’
calo / galo	[k]alo / [g]alo	‘corn’ / ‘cock’
gato / gado	ga[t]o / ga[d]o	‘’cat’ / cattle’

Table 2: Minimal pairs in EP

The dialectal variation of the following voiced fricative [z] is shown in Table 3 below.

Written forms	European Portuguese	Mozambican Portuguese	English
bazar	ba[z]ar	/b/a[z]ar; /p/a[s]ar /p/a[z]ar ; /p/a[ʃ]ar	‘bazaar’
prezado	pre[z]ado	pre[z]ado; pre[s]ado pre[ʃ]ado; pre[ʒ]ado	‘dear’

Table 3: Dialectal variation of voiced fricative [z] in some MP varieties

Table 4 shows the free variation of flap [ɾ][1] andalveolar trill [r][2] or uvular fricative [ʁ][3].

Written forms	European Portuguese	Mozambican Portuguese	English
areia	a[ɾ]eia	a[ɾ]eia; a[r]eia; a[ʁ]eia	‘sand’
carro	ca[ʁ]o; ca[r]o	ca[ɾ]o; ca[r]o; ca[ʁ]o;	‘car’
rádio	[ʁ]ádio, [r]ádio	[ɾ]ádio; [r]ádio; [ʁ]ádio,	‘radio’

Table 4: Free variation of flap [ɾ] andalveolar trill [r] or uvular fricative [ʁ] in MP varieties

[1] Also called “vibrante alveolar líquida” ‘liquid vibrant alveolar’ (Mateus et al., 1983; and Martins, M. R. D., 1992).

[2] Also called “vibrante múltipla” ‘multiple vibrant’ (Mateus et al., 1983; and Martins, M. R. D., 1992).

[3] Also called “vibrante velar” ‘velar vibrant’ (Mateus et al., 1983; and Martins, M. R. D., 1992).

The change and/or alternation of lateral [l] / flap [ɾ] is demonstrated in Table 5.

Written forms	European Portuguese	Mozambican Portuguese	English
falta	fa[l]ta	fa[l]ta / fa[ɾ]ta	‘lack’

Table 5: Change and/or alternation of lateral [l] and flap [ɾ] in MP varieties

Table 6 below exemplifies the chase of flap [ɾ] to lateral [l].

Written forms	European Portuguese	Mozambican Portuguese	English
garfo	ga[ɾ]fo	ga[ɾ]fo / ga[l]fo	‘fork’

Table 6: Chase of flap [ɾ] to lateral [l]

The lateralization or fricatization of lateral palatal [ʎ] is demonstrated in Table 7.

Written forms	European Portuguese	Mozambican Portuguese	English
mulher	mu[ʎ]er	mu[ʎ]er ~ mu[l]er ~ mu[ɬ]er	‘woman’

Table 7: Lateralization [l] or fricatization [ɬ] of lateral palatal [ʎ]

The rhotacisation [ɽ] of lateral [l] in a certain groups of consonants is indicated in Table 8 below.

Written forms	European Portuguese	Mozambican Portuguese	English
república	república	repúb[ɽ]ica	‘republic’
assembleia	assembleia	assemb[ɽ]eia	‘assembly’

Table 8: Rhotacisation [ɽ] of lateral [l] in groups of consonants

This study describes and discusses the following phonetic and phonological processes in Portuguese varieties of Mozambique:

(i) The complementary distribution vs free distribution of the phoneme /ʀ/ and

(ii) The lateralization and/or fricatization of the phoneme /ʎ/

1.4 Problem description

It is the aim of this thesis to investigate how the variants of each variable above are distributed. It is anticipated that there will be structural constraints as well as sociolinguistic factors pertaining to speakers and their background, including home language background. In order to achieve this goal it is essential to know the distinction between phoneme and its allophones. For example, in European and Brazilian Portuguese the phoneme /r/ (of *roda* wheel) can be pronounced with varying degree of vibration in a free variation, either as uvular fricative [ʁ] or as a trill alveolar [r]. All these different pronunciations constitute allophones of a single phoneme.

Allophones are phonetic variants of a phoneme in a particular language. Although long contested in traditional structural and descriptive phonological theory and overtaken by developments in generative phonology and metrical phonology, the phoneme-allophone relationship remains important for descriptive linguistics and sociolinguistics. Some examples are given below.

European Portuguese

- [ʁ] and [r] are allophones of the phoneme /ʀ/. Their relationship is discussed below.

Brazilian Portuguese

- [l] and [w] are allophones of the phoneme /l/. These allophones of the phoneme [l] are contextually motivated. Therefore, if they occur at the end of a syllable or word has different realization from that where it appears at the beginning of a syllable or word.

English

- [p] and [p^h] are allophones of the phoneme /p/.
- [t] and [t^h] are allophones of the phoneme /t/.

Note that in the European Portuguese varieties the realization of the phoneme /ʀ/ in the word *areia*, sand, (see example 3, above.) is flapped, i.e., instead of vibrating two or more times in a series of flaps, as in a normal trill, the speech organ vibrates just once. However, in the word *radio*, the realization of /ʀ/ is trilled, i.e., the breath stream and manner of articulation causes one of the vocal organs to vibrate many times. The flapped and trilled realizations are both recognized as /ʀ/ by European Portuguese speakers despite their differences. Nevertheless, flapped realization will never be found in the position in a word in which the trilled realization is appropriate, and vice versa. When we find this strict separation of environment in which particular realizations can occur, we say that the realizations are contextually motivated; that is, they are in complementary distribution (Roach 2000:41; Moulton 1962:5).

Allophones in a complementary distribution represent two different surface variants of the same phoneme, and the selection of one rather than the other is entirely and unambiguously dependent on the context. In neutralization processes, on the other hand, two contrasting phonemes merge in a specific context, in which one segment becomes the exclusive output of the allophonic rule (or at least, the most frequent variant), the other becomes rare, stigmatised or non-existent. The implications of the lack of control of the phonological rules of European Portuguese can be raised for African Mozambican speakers as L1 or as L2. I suggest that this situation accounts for what Stroud and Gonçalves (1997:3) called the "linguistic uncertainty" of many Mozambican learners and their major tendency to softening the phoneme /ʀ/ as an alveolar trill.

In the process of L2 phonological acquisition, L1 experience influences perceptual processing of speech sounds. It is not only a matter of presence vs. absence of phonological opposition; rather, there are different degrees of contrast and each may have a specific perceptual consequence. In fact, in L1 and L2 phonological acquisition the distinction between phonemes and allophones starts from the segmentation of the continuous speech

signal into discrete categories and the observation of their roles in the lexicon. Therefore, the core of perceptual acquisition (Harnad 1987; Gerrits 2001; Gerrits & Schouten 2004) is a process of sound discrimination, which depends largely on the degree of categorical perception achieved by the learner.

1.5 The phonological variables

My observations and experience of language usage were crucial to the selection of the relevant variables of this study. These variables are involved in a process of language variation and change of Portuguese. The size of the corpus of this study of sixteen subjects, each with at least twenty minutes of speech, enabled me to have a reliable picture of the phenomenon of language variation being studied. In chapter 3 I discuss in detail the selection process of variables and some technical constraints. To represent the variables I partially use the model proposed by Tarallo (1997:8-9) where vertical parallel bars / / indicate phonic (phonological) segments, i.e., the linguistic variables and the square brackets [ ] are used to indicate the variants. All sounds in round brackets denote rare occurrence in all or in some varieties.

Tables 9-12 summarize the linguistic variables and their variants selected for this study.

Variable	Variants	Phonetic futures	context
/ʀ/	[ʁ]	Uvular fricative	Initial / Middle
	[r]	Alveolar trill	Intial / middle
	[ɾ]	Tap or alveolar flap	Middle

Table 9: Phoneme /ʀ/and its EP variants

Variable	Variants	Phonetic futures	context
/ʀ/	([ʁ])	Uvular fricative	Initial / Middle
	[r]	Alveolar trill	Intial / middle
	[ɾ]	Tap or alveolar flap	Middle / Initial

Table 10: Phoneme /ʀ/and its MP variants

Variable	Variants	Phonetic futures	context
/ʎ/	[ʎ] EP	Lateral approximant palatal	Middle

Table 11: Phoneme /ʎ/ and its EP variants

Variable	Variants	Phonetic futures	context
/ʎ/	([ʎ]) MP	Lateral approximant palatal	Middle
	[l] MP	Lateral approximant alveolar	
	[ɬ] MP	Lateral fricative palatal	

Table 12: Phoneme /ʎ/ and its MP variants

The next chapter deals with the sociolinguistic context in which this study takes place. It describes the linguistic situation of Mozambique with special emphasis on the sociolinguistic context of Maputo by mother tongue (L1) whether Bantu or Portuguese, and whether Portuguese is a second language (L2) for the speaker.

1.6 The structure of the dissertation

This dissertation comprises six chapters.

Chapter 1

Aims to illustrate briefly some of the developments on acquisition research as the heart of modern study of language. It gives an overview of language contact and variation as fields of sociolinguistics. Further, this chapter also describes the linguistic situation of Mozambique. Special emphasis is on the sociolinguistic context of Maputo by mother tongue (L1) Bantu or Portuguese and Portuguese as a second language (L2).The main assumption is that any natural language is used in everyday life to satisfy communication needs within a specific geographical space and to signal personal and group identities. The chapter ends with a formulation of the problem and the research aim plus, and a statement of the hypothesis.

Chapter 2

Is dedicated to the literature review. Since language is a dynamic entity and as such, it is subject to variation and change, this chapter focuses on different studies and theories as well as research results on language variation and contact. It also analyses the most relevant studies on Portuguese sociolinguistics with emphasis to the context of language contact at the individual level in City of Maputo where Portuguese is in permanent contact with different Mozambican Bantu languages.

Chapter 3

Deals with methodology. First, this chapter gives a clear formulation of the research question. Also focuses on the process of elicitation of data with reference to the source of the data for this study, the selection of participants and collection of data. The main source of data for this study is the Maputo linguistic database known as *"Panorama do Português Oral do Maputo"* (henceforth PPOM), a linguistic survey gathered in the context of the study of oral Portuguese in Maputo published in 1997.

Chapter 4

Is related to the procedures and constraints. The corpus of this study was coded for a multivariate statistical analysis of VARBRUL. This chapter summarizes this quantitative analysis model, run using GoldVarb 2001 for Windows.

Chapter 5

Incorporates the presentation, discussion and the interpretation of the results of this study. The results were obtained from the tested linguistic variables in correlation with social factors using a multivariate function of GoldVarb 2001.

Chapter 6

Provides a summary of the conclusions, remarks and suggestions of directions to further studies. Findings confirm Chambers' (1998) principle 4 that acquisition of complex rules and new phonological features split into age groups indicating that alveolarization / uvularization of phoneme /R/ is a current change in progress among generations in Mozambican Portuguese. In fact, contrary to the European norm, in many varieties of the Mozambican variety of Portuguese the trill [r] (also called *vibrante alveolar múltipla sonora* (Martins,

Maria R. Delgado 1992:171 and Mateus et al. 1983:529); *erre multiple* or voiced alveolar trill (Morgan, Terrel 2006:119), or as uvular fricative [ʁ] (*vibrante velar múltipla Sonora* (Martins 1992 and Mateus et al. 1983), and the tap or alveolar flap [ɾ/] (called *vibrante alveolar simples sonora,* by Martins 1992 and Mateu et al. 1983), tendentially, occur in a free variation. The results of my study show that all the four social factors tested, namely residence, education, sex and age have significant effect on the sociolinguistic patterns displayed by all the sociolinguistic variables analysed in this study.

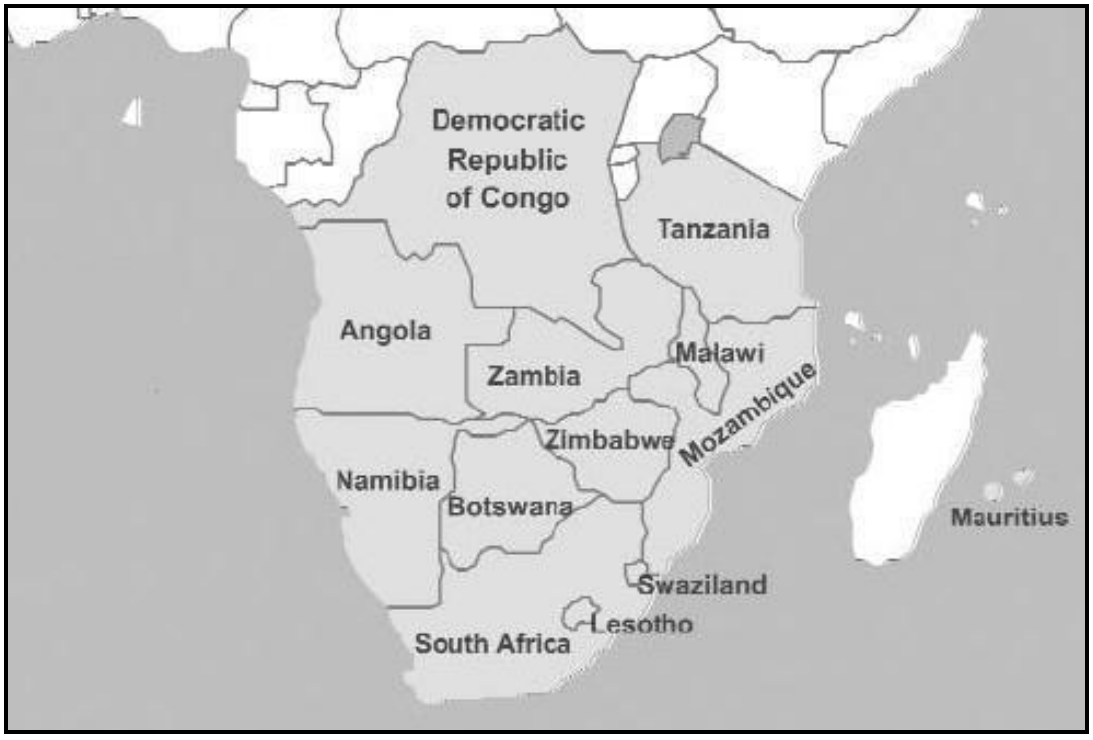
CHAPTER 2
LINGUISTIC SITUATION OF MOZAMBIQUE

Members of a specific community within a geographical space use any natural language in everyday life to satisfy daily communication needs. A language also can be used to differentiate its speakers in terms of their region, and to stratify them in terms of possible social factors like age, sex, ethnic-group and occupation.

This chapter is organized as follows: in section 1 I describe the linguistic situation of Mozambique. Section 2 focuses on the sociolinguistic context of the Maputo by mother tongue (L1) Bantu or Portuguese and Portuguese as a second language (L2). In section 3 I conclude with a description of the Bantu languages in contact situation with Portuguese in the selected area for this study.

2.1 Geographic situation of Mozambique

Mozambique is situated in the southern part of Africa. With a total surface of 799.380 Km^2, Mozambique is limited in north by Tanzania, in west by Malawi, Zambia, Zimbabwe and South Africa, and in south by Swaziland and South Africa. Mozambique has a cost of 2.515 Km along the Indian Ocean. The following Map 3 of Africa shows Mozambique within the Southern African Development Community region (SADC).



Map 3: Africa with SADC region

(Source: Institute of Security Studies, 2005)

2.2 Languages of Mozambique

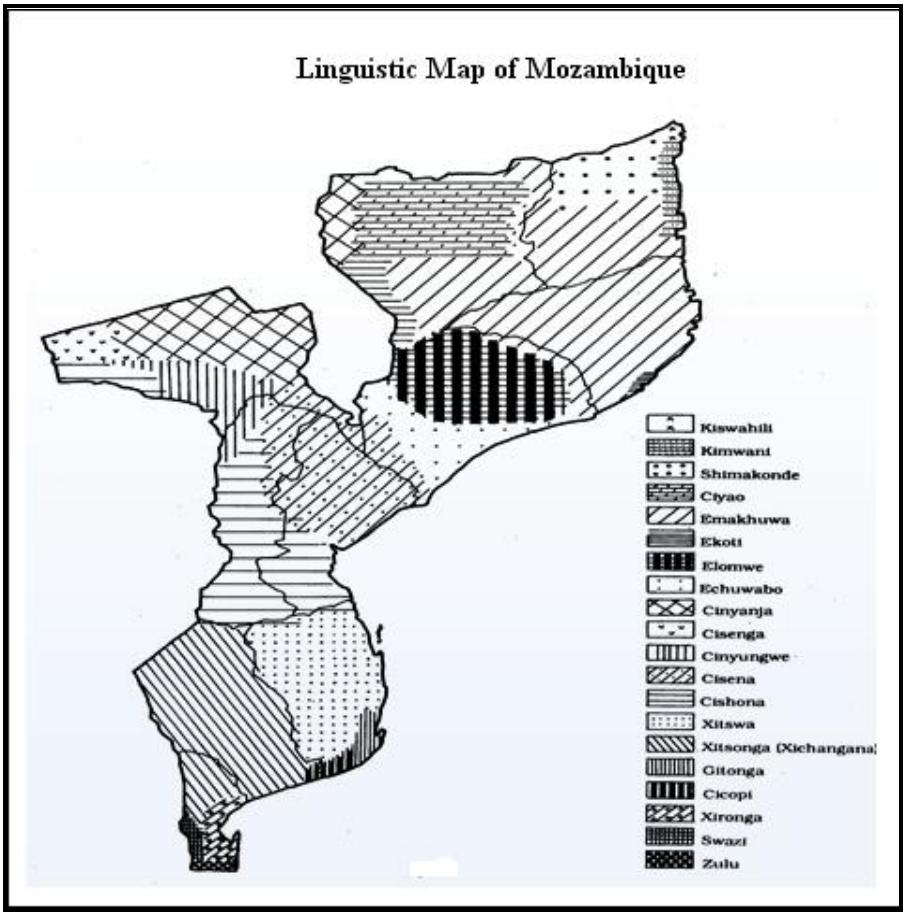
Following Guthrie's (1967-71) classification, with some exceptions of language pockets, mainly in South Africa, Botswana, Namibia, Kenya, and Tanzania, the communities in the region of contemporary Africa south, between the lines from the Cameroon Mountains, south of Nigeria, along the Atlantic coast, across the Central African Republic to southern Somalia in the east and the southern tip of Africa are Bantu speaking (Ngunga 2004:29-30). In a revised classification, Guthrie (1967-71) combining geographic and genealogical relationships grouped the Bantu languages into 15 linguistic zones represented by capital letters, namely, A, B, C, D, E, F, G, H, K, L, M, N, P, R and S. Each zone comprises a number of linguistic groups codified by decimal numbers ending by zero (0). The following numbers indicate the languages of the group. For example, in the following linguistic code S.50. Tswa-Ronga Group, the letter S represents the linguistic zone and the number 50 the group Tswa-Ronga. In its turn, S.53: Xichangani, after Guthrie, means that the language Xichangani is codified as number 3 and belongs to group 50 of the linguistic zone S (cited by Ngunga 2004:43-46).

Based on Ngunga (2000:43-49) and Sitoe and Ngunga (2000:25-217), in a revised version of NELIMO (1988), in a combination with the classification proposed by Guthrie (1967-71), the languages of Mozambique are as follows in Table 13:

Zones	Groups	Languages
Zone G:	G40: Group Swahili:	G.45: Kimwani
Zone P:	P.20: Group Yao:	P.21: Ciyao
		P.23: Shimakonde
	P.30: Makhuwa-Lomwe Group:	P31: Emakuwa
		P34: Echuwabu
Zone N:	N.30: Nyanja Group:	N.31 a: Cinyanja
	N.40: Nsenga-Sena Group:	N.43: Cinyungwe
Zone S:	S.10: Shona Group:	S.13 a: Cimanyika
		S.13 b: Ciutee
		Cibalke
		S.15 a: Cindau
	S.50: Tswa-Ronga Group:	S.51: Citshwa
		S. 53: Xichangana
		S.54: Xirhonga
	S.60: Copi Group:	S.61: Cicopi

Table 13: Mozambican languages' classification (Guthrie's (1967-71 in Ngunga 2004:46-48)

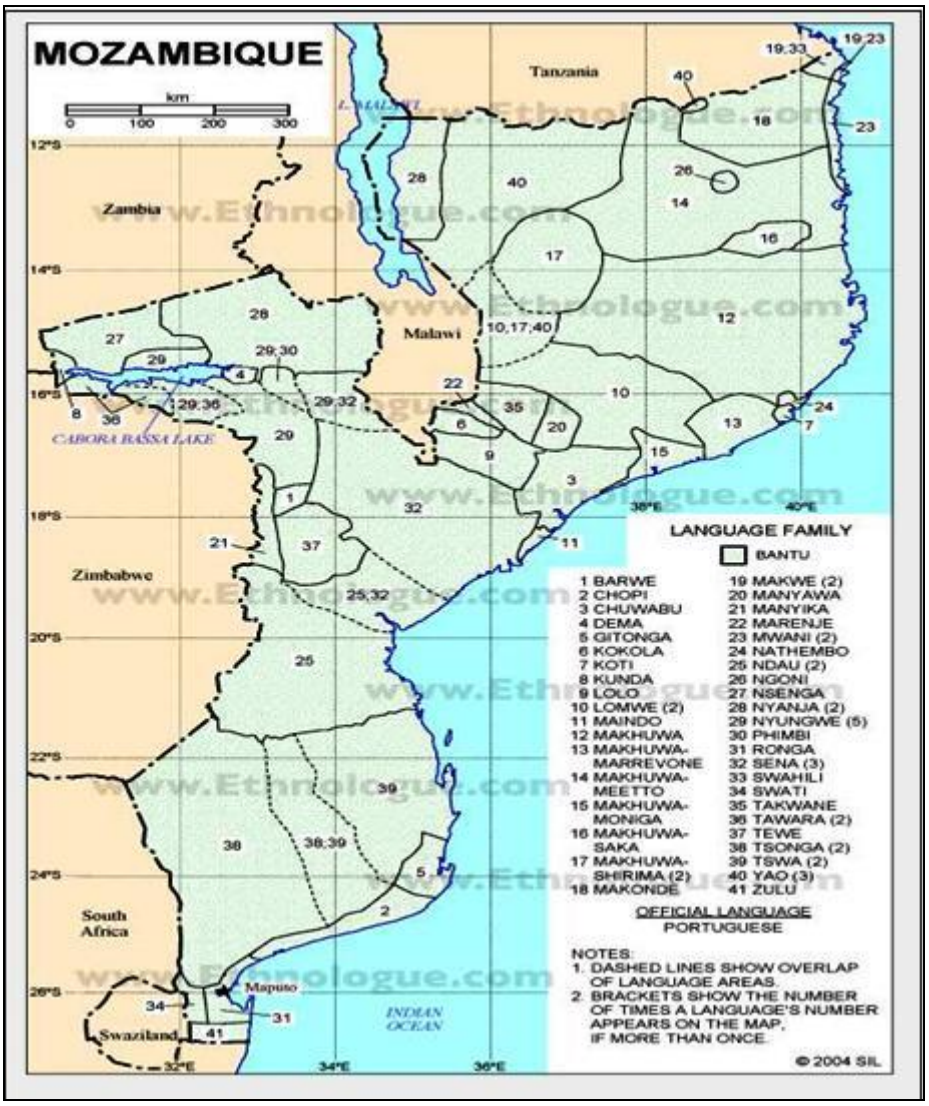
From Guthrie's inventory, only four linguistic zones (G, P, N, S) and eight groups and fifteen languages cover Mozambique. Map 4 below illustrates the linguistic density of Mozambique according to the results of the General Population Census in Mozambique in 1980. Light colours indicate the lowest density and the highest density is represented by the dark regions.



Map 4: Linguistic map of Mozambique (*NELIMO* 1989)

(Source: http://www.panafrill0n.org/wikidoc/pmwiki.php/PanAfrLoc/Mozambique)

There is no consensus yet about the exact names and the number of some of the languages spoken in Mozambique. Contrary to the above information, for example, where clearly are indicated 20 languages, map 5, from the *Ethnologue*, indicates a total of 40 languages. Besides the possibility of the emergence of new languages, I believe that partially this discrepancy of numbers clearly results from the criteria that are used. The case of Citshwa, Xichangana and Xirhonga, mutually intelligible but considered as separate languages, is illustrative. This situation is motivated by a strong separation of social and cultural functions of these three linguistic entities. In fact, the speakers of these three varieties understand each other at a higher level, almost 80%, given the fact that they use common vocabulary, syntactic constructions and meaning. Like this situation, there are many examples in different regions of Mozambique. Therefore, such cases of higher linguistic diversity like this represented by Map 5 below, from the *Ethnologue*, deserve further comments in future researches.



Map 5: Linguistic map of Mozambique (SIL 2004)

(http://www.ethnologue.com/show_map.asp?name=MZ&seq=10)

Note also that the linguistic inventories of map 4 (*NELIMO* 1988), and map 5 (SIL 2004, in *Ethnologue*), include languages spoken in Mozambique by minorities along the borders with South Africa, the Zulu, with Swaziland, the Swati, and with Tanzania, the Swahili. However, Sitoe and Ngunga (2000:25-217) and Ngunga (2000:43-49) don't include these languages their lists.

2.3 (Some) Phonological and phonetic comparative elements between the European Portuguese and the Mozambican Bantu languages

Phonetics is primarily concerned with the description of how the sounds of the languages of the world are made, i.e., which aspects of the sound are essential to convey meaning (Ladefoged 1993:1). Phonetics, therefore, deals with sounds as physical products in relation to the properties of human beings that produce them (Lass 1984:1). On the other hand, the immediate concern of phonology as a sub-discipline within linguistics is to study the "function, behaviour, and organisation of sounds as linguistic items" (Lass 1984:1). This section of this chapter is to present a general outline of the phonetic and phonological characteristics of EP and Mozambican Bantu languages. From the phonetic and phonological point of view is to show that between the EP and Mozambican Bantu languages there are substantial differences, which can affect upon the ongoing variation in the MP, resulting from the language contact between them. In fact, although theoretically at the abstract level, i.e., the phonological level, MP and EP use the same inventories there are variations resulting from the individual production of certain sounds that can be interpreted by the speakers of the two varieties of Portuguese as belonging to different elements of the system. According to Mateus et al. (1983:526) EP's contain the following consonantal sounds:

(i) Consonants at the initial position: [p], [b], [t], [d], [k], [k], [f], [v], [s], [z], [ʃ], [ʒ], [l], [m], [n], [ʀ];

(ii) All consonants in (1) can occur the middle position, plus: [ɲ], [ʎ], and [r];

(iii) At the end of the syllable (i.e., before another consonant at the beginning of the following syllable) and at the end of the word only the following consonants occur: [l], [r], and [ʃ].

Although these authors do not mention specifically the case of MP, they refer to the existence of different norms and dialects of Portuguese with some sounds that are not included in their list. Besides this fact it is important also to notice that there are dialectal varieties of MP without some of the sounds included in the above inventory, for example [b],

[d], and [g]. This case can be one of the consequences of language contact between Portuguese and Bantu languages without the referred sounds in their phonological systems, such as in one of Guthrie's (1967-71, cited by Ngunga 2000:43-49) Zone P Languages, the Emakhuwa. The following Table 14 represents a revised standard orthography. The first table was proposed by *NELIMO* (1989), revised by Sitoe and Ngunga (2000) and by Ngunga (2004). Note that according to the authors cited above all sounds represented by symbols in round brackets are rare in the referred language.

IPA	Mwa	Mak	Yao	Ma	Chu	Nya	Nyu	Sen	Man	Nda	Ute	Bal	Ton	Cop	Cha	Tsh	Rho
b	b	b	b	(b)	b	(bh)	bh	bh	bh	bh	bh	bh	bh	bh	b	bh	b
ɓ	-	-	-	-	-	b	b	b	b	b	b	b	b	b	b'	b'	b'
d	d	d	d	(d)	d	(dh)	dh	dh	dh	dh	dh	dh	dh	dh	d	dh	d
ɗ	-	-	-	-	-	d	d	d	d	d	d	d	d	d	d'	d	d'
ð	-	-	-	-	dh	-	-	-	-	-	-	-	-	-	-	-	-
g	g	g	g	(g)	g	g	g	g	g	g	g	g	gh	g	g	g	g
ɣ	-	-	-	-	-	-	-	-	-	-	-	-	g	-	-	-	-
c	c	c	c	c	c	c	c	c	c	c	c	c	c	c	c	c	c
dʒ	j	j	j	(j)	j	j	dj	dj/j	j	j	j	j	-	j	j	j	j
ʃ	sh	sh	-	x/sh	-	sh	x	x	sh	sh	sh	sh	-	x	x	x	x
v	-	v	-	v	v	vh	(vh)	(vh)	vh	vh	v	-	vh	vh	vh	vh	vh
ʋ	-	-	w/v	-	-	v	v	v	v	v	v	-	v	v	v	v	v
β	-	-	-	-	-	-	-	-	-	-	-	-	vb	-	-	-	-
ʃv	-	-	-	-	-	-	sw	sw	sv	sv	sv	sv	-	(sv)	sv	sv	sv
ʒv	-	-	-	-	-	-	zv	zw	zv	zv	zv	zv	-	(zv)	zv	zv	zv
pʃ	-	-	-	-	-	ps	ps	ps	-	-	-	ps	-	ps	ps	ps	ps
bʒ	-	-	-	-	-	bz	bz	bz	-	-	-	bz	-	bz	bz	bz	bz
tʃ	-	-	-	-	-	-	-	-	tsv	tsv	tsv	-	-	-	-	-	-
dʒ	-	-	-	-	-	-	-	-	dzv	dzv	dzv	-	-	-	-	-	-
ŋ	ng'	ng'	ng'	ng'	ng'	ng'	ng'	ng'	n'	n'	n'	n'	n'	n'	n'	n'	n'
ɴ'	N'	N'	N'	N'	N'	N'	N'	N'	-	-	-	N'	-	m'/m	-	-	-
l	l	l	l	l	l	l	l/r	l/r	-	-	-	l/r	l	l	l	l	l
ɬ	-	-	-	-	-	-	-	-	-	-	-	-	-	-	hl	hl	hl
ʎ	-	ly/dy	-	ly	dy	dy	dy	dy	-	-	-	dy	-	lh	lh	-	lh
r	-	-	(r)	r	r	-	-	-	r	r	r	-	r	r	r	r	r
ɾ	-	-	-	-	lr	-	-	-	lr	-	-	-	-	-	-	-	-
ɽh	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	rh

Table 14: Standard Orthography of the Mozambican Languages
(My version from *NELIMO* 1989, Sitoe and Ngunga 2000 and Ngunga 2004)

Table 15 below shows a descriptive summary of the variable sounds of this study, namely the lateral and vibrant consonants of EP and of the African Bantu languages spoken in Mozambique.

			Alveolar	Palatal	Uvular	
Consonants		**LATERAL**				**Changana**
		Voiceless		hl		
		voiced	l, dl	lh		
		VIBRANT	r		-	
		LATERAL				**Ronga**
		Voiceless		hl		
		voiced	l, dl	lh		
		VIBRANT	r		-	
		LATERAL				**Tonga**
		Voiceless		-		
		voiced	l	-		
		VIBRANT	r			
		LATERAL				**Macua**
		Voiceless		-		
		voiced	l	ly		
		VIBRANT	r		-	
		LATERAL				**Sena**
		Voiceless		-		
		voiced	l/r	-		
		VIBRANT	r/l		-	
		LATERAL				**Portuguese**
		Voiceless		-		
		voiced	l	lh		
		VIBRANT		r	ʁ[1]	

Table 15: European Portuguese vs Mozambican Bantu languages lateral and vibrant consonants ***(NELIMO*** **1989, Sitoe and Ngunga (2000:69, 106, 157, 177, 207); Ngunga (2004), Mateus et al. (1983:529-530)**

[1] It also occurs as alveolar trill in some dialects and is represented by [r].

For the purpose of the present study, the sounds referred to in the table above as vibrant correspond to the trill and flap or tap sounds according to the International Phonetic Alphabet (cited by Martins 1998:168). Comparatively, the table above shows three important differences between Portuguese and the comparative African Bantu languages: (1) the absence of the vibrant velar consonant in all African Bantu languages, (2) the absence of the consonant lateral voiced palatal in two of the comparative African Bantu languages, and 3) the occurrence of a consonant lateral voiceless palatal in three of the comparative African Bantu languages. In my point of view these three aspects, individually or combined, could be very important to explain some of the sociolinguistic phenomena analyzed in this study.

Portuguese is the only declared official language and is no longer a minority language. In effect, according to the III General Census held in 2007, in the city of Maputo with 1,099,102 inhabitants, approximately 43% of its population aged 5 or more speak Portuguese as mother tongue. Note that with this figures Maputo is the major centre where Portuguese is spoken as a mother tongue in Mozambique.

According to the 1997 Mozambican second census the population of Mozambique was estimated in 16,099,246 inhabitants, 8.8% of which speaks Portuguese as a first language and 27% as a second language. Regarding to the Bantu languages the census showed the following figures: eMakhuwa 26.1%, xiChangana 11.3%, eLomwe 7.6%, ciSena 6.8%, and eChuwabo 5.8 of the whole population[1]. The most recent findings after the Third Population and Housing Census held in 2007 reported an estimated population of 21,669,278 inhabitants in 2009. Approximately 40% of the national population speaks Portuguese. In urban areas the percentage of Portuguese speakers is estimated in about 72.4% of inhabitants. (See INE - Mozambique's National Institute for Statistics 2009 cited by *CIA - The World Factbook*).

[1] See: http://www.historycentral.com/nationbynation/Mozambique/Index.html

2.4 Population of Mozambique

According to the 1997 Second General Population and Housing Census (*INE*, 1999) the population of Mozambique was 16,1 million. In 2007 it increased to 20,5 million, according to the Third General Population and Housing Census (INE, 2008). Although the majority of the population still live in rural areas, there is a continuous increasing of the urban population. Among others there are two key factors for this growing, namely, (i) the effect of the armed conflict between the Mozambican National Resistance (RENAMO), now transformed in a political opposition party, and Frelimo, the ruling party, and (ii) the local migrations motivated by the attractive basic urban expansion along the country, especially in city of Maputo and Maputo province.

Indeed, at the beginning of the 1980s, 13 percent of the population were located in urban centres, but by mid-1990s this figure significantly increased to 27 percent (Hatton et al. 2001:21). The figures below show comparatively the growth of the total population of Mozambique between the last two general population and housing censuses. This rapid increase of the population has also linguistic implications, namely the increase of the Portuguese speakers' population via school, that is, the literate people, more in terms of acquired skills than in the perspective of skills usage. The growth of the total population of Mozambique by provinces in 1997 and 2007 can be seen in Table 16.

Provinces	Population		Variation
	1997	2007	1997-2007
Total	**16,075,708**	**20,530,714**	**27.7**
Niassa	808,572	1,178,117	45.7
Cabo Delgado	1,380,202	1,632,809	18.3
Nampula	3,063,456	4,076,642	33.1
Zambézia	3,096,400	3,892,854	25.7
Tete	1,226,008	1,832,339	49.5
Manica	1,039,463	1,418,927	36.5
Sofala	1,368,671	1,654,163	20.9
Inhambane	1,157,182	1,267,035	9.5
Gaza	1,116,903	1,219,013	9.1
Maputo Province	830,908	1,259,713	51.6
City of Maputo	**987,943**	**1,099,102**	**11.3**

Table 16: The Growth of the Total Population, by Provinces 1997 and 2007
(Source: *Instituto Nacional de Estatística (INE).*

2.5 Population distribution of City of Maputo by age, sex and mother tongue

The following section shows the distribution of the population of Maputo within the selected age brackets for this study, according to the results of the Third General Population and Housing Census in Mozambique in 2007. Tables 17, 18, and 19 (Source: *Instituto Nacional de Estatística - INE*) show the total population of the City of Maputo by age, according to mother tongue, including the Portuguese language and the main African Bantu languages, namely, Xichangana, Xirhonga, Citshwa e Gitonga. Not that the population included in the tables is only that situated within the age brackets of my analysis.

AGE BRACKETS 10–34: YOUNG AGE

				10-14	15 - 19	20 - 24	25 - 29	30 - 34
Maputo City	2007	HM(Total)	Total	128 570	125 846	135 739	112 039	76 446
			Portuguese	75 000	64 702	61 699	46 167	25 670
			Xichangana	35 963	38 678	42 978	35 110	25 841
			Xirhonga	10 135	10 360	11 078	9 878	7 987
			Cicopi / Cichopi	1 371	2 501	3 444	3 719	3 235
			Citshwa/ Xitshwa	1 446	3 091	5 058	4 795	3 667
			Gitonga / Bitonga	953	1 861	3 119	3 261	2 887
			Other Mozambican languages	1 972	3 145	6 315	7 161	5 462
		Males	Total	63 024	62 233	65 397	55 226	36 014
			Portuguese	36 927	31 621	29 915	22 881	12 402
			Xichangana	17 690	19 221	19 624	15 815	10 693
			Xirhonga	5 036	5 242	5 182	4 540	3 332
			Cicopi / Cichopi	562	1 141	1 546	1 711	1 504
			Citshwa / Xitshwa	650	1 666	2 484	2 399	1 773
			Gitonga / Bitonga	412	903	1 395	1 544	1 434
			Other Mozambican languages	872	1 740	4 253	5 230	3 897
		Females	Total	65 546	63 613	70 342	56 813	40 432
			Portuguese	38 073	33 081	31 784	23 286	13 268
			Xichangana	18 273	19 457	23 354	19 295	15 148
			Xirhonga	5 099	5 118	5 896	5 338	4 655
			Cicop i/ Cichopi	809	1 360	1 898	2 008	1 731
			Citshwa / Xitshwa	796	1 425	2 574	2 396	1 894
			Gitonga / Bitonga	541	958	1 724	1 717	1 453
			Other Mozambican languages	1 100	1 405	2 062	1 931	1 565

Table 17: Population of Maputo city by gender and mother tongue in the age brackets of 10-34

However, is important to notice that, besides these languages, there are other non-indigenous languages spoken in the city by a very significant number. Indeed, according to the final 2007 census' results, the total number of speakers of such other African Bantu languages is even superior to that of Xirhonga and Citshwa, individually.

Table 18 below illustrates the distribution of the population in Maputo within the age brackets of 35-54.

AGE BRACKETS 35-54: MIDDLE AGE

				35 - 39	40 - 44	45 - 49	50 - 54
Maputo City	2007	HM(Total)	Total	58 965	51 106	44 734	33 187
			Portuguese	15 769	11 923	9 351	6 614
			Xichangana	20 161	17 980	16 931	12 683
			Xirhonga	7 236	6 319	5 785	4 806
			Cicopi / Cichopi	3 272	3 369	3 058	2 277
			Citshwa / Xitshwa	3 298	3 036	2 681	1 855
			Gitonga / Bitonga	2 855	2 700	2 526	2 060
			Other Mozambican languages	4 920	4 516	3 378	2 056
		Males	Total	26 591	23 758	22 212	16 777
			Portuguese	7 392	5 737	4 790	3 622
			Xichangana	7 698	7 064	7 446	5 736
			Xirhonga	2 701	2 372	2 328	2 005
			Cicopi / Cichopi	1 500	1 551	1 594	1 211
			Citshwa / Xitshwa	1 553	1 544	1 555	1 071
			Gitonga / Bitonga	1 336	1 330	1 312	1 127
			Other Mozambican languages	3 569	3 413	2 641	1 543
		Females	Total	32 374	27 348	22 522	16 410
			Portuguese	8 377	6 186	4 561	2 992
			Xichangana	12 463	10 916	9 485	6 947
			Xirhonga	4 535	3 947	3 457	2 801
			Cicopi / Cichopi	1 772	1 818	1 464	1 066
			Citshwa / Xitshwa	1 745	1 492	1 126	784
			Gitonga / Bitonga	1 519	1 370	1 214	933
			Other Mozambican	1 351	1 103	737	513

Table 18: Population of Maputo city by to gender and mother tongue in the age brackets of 35-54

AGE BRACKETS 55+: OLD AGE

				55 - 59	60 - 64	65 69	70-74	75- 79	80 +
Maputo City	2007	HM (Total)	Total	21 374	14 320	10 283	7 044	4 398	3 761
			Portuguese	3 850	2 164	1 459	928	612	516
			Xichangana	8 313	5 636	4 260	2 909	1 836	1 609
			Xirhonga	3 520	2 487	1 859	1 420	908	807
			Cicopi / Cichopi	1 564	1 083	814	603	328	267
			Citshwa / Xitshwa	1 178	825	539	271	188	145
			Gitonga / Bitonga	1 345	1 115	747	515	295	214
			Other Mozambican languages	995	577	352	206	116	107
		Males	Total	10 892	7 141	4 861	3 074	1 698	1 124
			Portuguese	2 274	1 262	816	479	315	216
			Xichangana	3 874	2 549	1 838	1 160	609	431
			Xirhonga	1 428	1 001	725	540	320	187
			Cicopi / Cichopi	786	572	411	279	141	80
			Citshwa / Xitshwa	683	479	309	131	70	46
			Gitonga / Bitonga	750	640	429	298	150	87
			Other Mozambican languages	742	395	216	118	53	48
		Females	Total	10 482	7 179	5 422	3 970	2 700	2 637
			Portuguese	1 576	902	643	449	297	300
			Xichangana	4 439	3 087	2 422	1 749	1 227	1 178
			Xirhonga	2 092	1 486	1 134	880	588	620
			Cicopi / Cichopi	778	511	403	324	187	187
			Citshwa / Xitshwa	495	346	230	140	118	99
			Gitonga / Bitonga	595	475	318	217	145	127
			Other Mozambican languages	253	182	136	88	63	59

Table 19: Population of Maputo city by gender and mother tongue in the age brackets of 55 and more

According to the III population census held in 2007 the majority of the Mozambican population is very young. To illustrate this fact, in 2003, Mozambique with a population estimated by the United Nations[1] at 18,863,000, only 3% of it was over 65 years of age, with another 45% of the population under 15 years of age. The III Population Census (2007) indicates that in the city of Maputo 36.1% of the population is situated within the age brackets of 0-14 years while 60.3% is within the age brackets of 15-59. Only 3.6% of the population in the city of Maputo is within the age brackets of 60+.

[1] http://www.nationsencyclopedia.com/Africa/Mozambique-POPULATION.html#ixzz0YLKBVzBT

It is therefore important to take into account in my analysis of the data in chapter 5 the higher incidence of Portuguese speakers among young people. Also relevant from the tables is the fact that the major mother tongue in Maputo is the Portuguese language, followed by Xichangana on one hand, and on the other the fact that the majority of females in Maputo, comparatively to males, have Portuguese as their mother tongue, as per indicated in tables 6 and 7 above.

The following Table 20 is a summary of the population distributed by the urban districts studied (See http://www.ine.gov.mz/censo2007/rp/pop07/maputocidade and *INE* 1997). The city of Maputo is divided in administrative areas called *Distrito Urbano* (henceforth DU) codified by decimal numbers, from 1 to 7 indicating the number of the districts of Maputo. Each district comprises a certain number of residential areas. For this research I studied the population of the *DU1* (residential areas of Alto-Maé A and B, Polana Cimento A and B), *DU2* (residential areas of Chamanculo A, B, C and D), and *DU3* (residential areas of Maxaquene A, B, C and D and Mafalala).

Residential Area	Urban District
Polana Cimento "A" *Polana Cimento "B"* *Alato-Maé "A"* *Alato-Maé "B"*	*Distrito Ubrbano No. 1*
Chamanculo "A" *Chamanculo "B"* *Chamanculo "C"* *Chamanculo "D"*	*Distrito Urbano No. 2*
Maxaquene "A" *Maxaquene "B"* *Maxaquene "C"* *Maxaquene "D"* *Mafalala*	*Distrito Urbano No. 3*

Table 20: City of Maputo by urban districts and residential area

Note that although the general designation of the administrative areas of the city of Maputo gives the impression that this city comprises only urban areas, in fact Maputo also has suburban areas. In some cases a same administrative area is a continuum with an urban in one extreme and in the other a suburban area.

In Mozambique, by extension of the current use of the term in many countries of the developed world (see http://en.wikipedia.org/wiki/Suburban#Suburbia_worldwide), suburban areas are defined in opposition to urban areas, i.e., the central area of a major city, usually with high population density. However, economically and even in terms of delinquency rates or social problems in many cases it is not easy to draw a clear line between the inhabitants of most of suburban areas of Mozambican cities and/or between suburban and urban areas.

One of the social factors analyzed in this study and susceptible of having an significant impact on the levels of variation of Portuguese in Mozambique and especially in Maputo city is education. Tables 21-23 below show the distribution of the population in the city of Maputo within the age brackets of 5-10, 10-15, and 15-25 and+ by sex and the condition of attending a school in the City of Maputo (See *INE* - Mozambique's National Institute for Statistics, 2007).

				Total	Males	Females
City of Maputo	2007	Is attending a school	5	6 503	3 188	3 315
			6	21 748	10 704	11 044
			7	26 805	13 175	13 630
			8	25 169	12 273	12 896
			9	24 134	11 697	12 437
			10	25 031	12 353	12 678
		Is not attending a school	5	113	52	61
			6	490	241	249
			7	446	211	235
			8	523	246	277
			9	472	225	247
			10	621	291	330
		Never attended a school	5	17 680	8 953	8 727
			6	4 479	2 285	2 194
			7	1 675	829	846
			8	808	410	398
			9	615	312	303
			10	503	245	258

Table 21: Attendance of school - age brackets 5-10

				Total	Males	Females
City of Maputo	2007	Is attending a school	10	25 031	12 353	12 678
			11	24 096	11 905	12 191
			12	25 841	12 792	13 049
			13	24 294	11 742	12 552
			14	22 772	11 121	11 651
			15	22 081	10 976	11 105
		Is not attending a school	10	621	291	330
			11	646	313	333
			12	793	378	415
			13	938	448	490
			14	1 321	625	696
		Never attended a school	10	503	245	258
			11	514	248	266
			12	424	195	229
			13	381	172	209
			14	395	196	199
			15	541	251	290

Table 22: Attendance of school - age brackets 10-15

				Total	Males	Females
City of Maputo	2007	Is attending a school	15	22 081	10 976	11 105
			16	20 539	10 314	10 225
			17	19 035	9 865	9 170
			18	17 246	9 098	8 148
			19	13 836	7 375	6 461
			20-24	48 629	24 719	23 910
			25 e+	47 290	21 734	25 556
		Is not attending a school	15	2 154	1 008	1 146
			16	3 637	1 595	2 042
			17	5 781	2 481	3 300
			18	8 525	3 862	4 663
			19	10 049	4 375	5 674
			20-24	80 833	38 555	42 278
			25 e+	345 379	177 998	167 381
		Never attended a school	15	541	251	290
			16	506	229	277
			17	519	228	291
			18	677	300	377
			19	720	276	444
			20-24	6 277	2 123	4 154
			25 e+	44 988	9 636	35 352

Table 23: Attendance of school - age brackets 15-25 and +

Very roughly the distribution of the population in education in the city of Maputo shown in tables 21-23 indicates a clear balanced situation between females and males attending a school, except for female over 25. This situation contrasts with the reality in the countryside where many studies refer to a higher disadvantage for the females compared to their male's pairs in terms of the levels of attendance of school. If education is one of the significant factors in language variation then the city of Maputo will play a great role in the possible change of PM in future given the fact that it is the major centers of diffusion of Portuguese in Mozambique.

CHAPTER 3
LITERATURE REVIEW

Traditionally, linguistic research has given privilege to the native speakers as its main source of language data neglecting the non-mother tongue users (Ferguson 1983:vii). This practice seems to be essentially influenced by Chomsky and generative-transformational tradition based on the 'system' view approach of language (Appel & Muysken 1987:153), according to which the linguist's task is the study of linguistic competence of the ideal listener-speaker in a linguistically homogenous speech community (Chomsky (1965, quoted by Terralo (1997:6)). On the other hand, is the 'bag of tricks' view, commonly associated with Hymes and other functionalists, that assigns to languages, as complex tools, the role in referring to the items and processes in communicating about them, due their easy capacity to adapt to new communicative and referential needs (Appel & Musken 1987:153).

The outline of this chapter is as follows: in section 1 I review the literature on theoretical approaches to the nature of sociolinguistic study, as related to social uses of language and to claims about the existence of a multitude of modes of inquiry. In section 2 I summarise three of the eight principles of dialect acquisition as postulated by Chambers (1998), with an attempt to show how they shed light on some of the problems of language acquisition. In section 3 I give an overview on the emergence of new varieties with special emphasis on the Portuguese language in Mozambique, the fieldwork of this study. I conclude in section 4 a discussion of second language acquisition and learning process and its implications in language variation.

3.1 The nature of sociolinguistic study as related to social uses of language

In the beginning of the last quarter of the twentieth century modern linguistic theory was marked by the consolidation of the point of view that any attempt to study speech must consider as its starting point the relationship between language and society. Or put differently, according to Labov (1972a: xiii), there cannot be a 'successful linguistic theory or practice which is not social'. By the light of this assumption and considering the sociolinguistic nature of his study, I decided to clarify briefly what should meant by the term

sociolinguistics. As Labov continues, 'the basis of intersubjective knowledge in linguistics must be found in speech', that is, the social uses of language. Indeed, as Chambers (1995:2) observed, the most relevant information about the person we are listening to, in terms of personal characteristics, background or character, is inferred in the very simple verbal exchange. He established five general categories about the nature of linguistic inferences, namely: personal, stylistic, social, sociocultural and sociological. In the following sub-sections I describe the five categories of inferences we make when we are engaged in a conversation.

3.2 The category of personal characteristics

The personal level, argues Chambers (1995:3), gives information about the voice quality and speaking ability of the individuals we are listening to. In effect, Johnson (1991:557), quoted by Chambers, drawing on her study of interviewers in Texas Telephone Poll, a standardized state-wide survey, says with regard to personal characteristics that 'conversation is always a kind of personal expression, (…)'. Therefore, Chambers (1995:4) concludes that although the information inferred from personal linguistic characteristics seems to be of little interest to some sociolinguists, it should be very relevant in the study about the way 'how personal speech characteristics vary across societies, or conversely, how they remain constant across social and cultural boundaries'. In this regard the stylistic category which is described next is important in sociolinguistic inquiry.

3.3 The category of linguistic styles

One instance of sociolinguistic relevance is stylistic variation. The stylistic level of observation allows the listeners to discern, among other aspects, the degree of intimacy or distance between participants in an interaction, their relative ages and social classes, and the function of their conversation, through the speech style they are exploring. According to Chambers (1995:5), the degree of formality is directly proportional to the number of differences between participants, viz., the greater the social differences the greater the formality in speech style and the most relevant social factors are the topic and age. The sociolinguistic importance of speech style derives from the fact that our ability to discern the formality of a conversation is largely determined by linguistic cues. Chambers (1995)

concludes that spontaneous talk is usually more rapid, with more syntactic ellipses and contractions, and more phonological assimilations and coalescences.

Conversely, although highly formal conversations can also be rapid, syntactically they tend to be more elaborated and phonologically 'conservative' in overriding casual phonology. Two cases of stereotypical hyper-formality examples can be mentioned here, namely:

(i) the hyper-correct [eɪ] of the English indefinite article "a" usually pronounced [ə] (mentioned by Chambers 1995:5),

(ii) and a tendency in MP variant among young people to velarization of [ʁ] in hyper-correct style of the tap or flap [ɾ], which is ordinarily pronounced as a flap in EP variant.

In this context, Chambers (1995 citing Fisher 1958:49 and Labov 1966:90-135), argues that, if the variation of formality in a conversation can cause speakers to adjust their speech norms, among other aspects of language, such as their phonology and accent, then style is an independent variable. This variable is a manipulated factor that affects the dependent speech variables, i.e., the observed and measured factors that vary according to the independent variable.

Note that all languages are susceptible to show variations in style, although the exact range of these variations and the level of linguistic structure at which they appear vary significantly (Ferguson and Gumperz 1973:98). In the following subsection, I will be particularly concerned with the nature of speech social characteristics and their relation with social stratification.

3.4 The category of social characteristics

As one of the most emblematic personal and social characteristics, speech shows the hallmarks of our social background. An individual's choice from among permissible alternates in a particular speech event may reveal family background and the social intent, the region and the social status, education, and may even indicate if the person "wishes to appear friendly or

distant, familiar or deferential, superior or inferior" (Gumperz 1972:220). In fact, our social class regulates our speech and it includes, among other aspects, education, occupation and type of housing. More specifically, he argues in this regard that the verbal interaction is a social process governed by social recognized norms and expectations, where the distribution of linguistic variants reflects social facts. Therefore, he concludes that the control of communicative resources varies in accordance with the individual's position within the social system and the need of verbal resources facility is directly proportional to the size of individual's sphere of activities, so that the more narrowly confined the sphere of activities, the more homogenous the social environment of interaction, and less need for verbal facility (Gumperz 1972:226).

In this context, the study of linguistic phenomena within a specific speech community is primarily concerned with language usage as it reflects general norms of this socially defined universe. Consequently, the communication of social information presupposes the existence of regular relationships between language usage and social structure, and social norms of language choice vary from situation to situation, from group to group and from community to community. Therefore, as Gumperz points out, the verbal markers of social distinctions have the same importance so that in multilingual societies the choice of one language over another has the same signification as the selection among lexical and phonetic or syntactic alternates in linguistically homogeneous societies (Gumperz 1972:229).

Although this determinism of social norms to set limits to freedom communication in highly stratified societies seems to be a consolidated fact, it is difficulty to generalise to all situations, such as the multilingual changing societies. Mozambique for example, where the flow of communication is dominated by single all-important urban centres, constituted by constellations of small highly active commercial conglomerates, shows "a range of immediate varieties bridging the transitions between extremes" (Gumperz 1972:230), in a fluid of repertoires so that is difficulty to draw a clear borderlines between different styles. Of course, in these contexts, social change causes both a weakening of traditional social structures and the formation of new ties and, for the first time, makes it possible the majority to free access to the privileged norm previously restricted to the minority assimilated group.

3.5 Language variation and change

Alternations in the mechanism of language system of sounds and grammatical categories have been the fundamental process of dialect and language changes over centuries (Labov 2001:6). In this section I will be focusing on the main trends related to language variation and change and some of the principles governing language change.

Although change is as inherent to language as it is in biological species it is difficult to accept it at all. Therefore and parallel to biological evolution, sound change has been seen negatively as it is considered a decay resulting in the loss of meaning (Labov 2001:10 citing Lehmann 1967:90). However, as Labov (2001:14) claims, even considering this parallelism and identity in the evolution of species and the evolution of languages, the fundamental causes of evolution in the two processes are totally different.

3.5.1 The causes of sound change: some trends

The interest and effort to identify and explain the causes of sound change go up to the 19th century. The magnitude of this difficulty in explaining the causes of sound change is clearly stated by Labov (2001:16), citing Bloomfield1933), as follows:

> Although many sound-changes shorten linguistic forms, simplify the phonetic systems, or in some other way lessen the labor of utterance, yet no student has succeeded in establishing a correlation between sound-change and any antecedent phenomenon: the causes of sound-change are unknown.

In spite of the number of explanations of linguistic change, my concern is only with those regarding language and dialect contact. Far from the pessimistic views of language change dominating the 19th and the early 20th centuries, recent approaches consider change as "maximizing the flow of information and ease of obtaining it" (Labov 2001:20, citing Martinet 1955). However, it was in the course of the first half of the 20th century where the shift took place with Sturtenvant's views that the path of linguistic change is a reflection of social change. In spite this achievement, Labov considers that the main concerns on who causes linguistic change still have to be accounted for (2001:29). Therefore, he argues, if

some speakers in the social architecture are responsible for language change then it becomes crucial to account for their social class, sex, ethnicity or their occupation.

Among those who looked on sound change with suspicion is Whitney to whom innovations in language were attributed to lowest social classes which was accused of being ignorant of the normative standard and negligent, as explicit on Labov's citation below:

> The whole regrettable inaccuracies of heedless speakers, their confusion of things which ought to be carefully held apart, their obliteration of valuable distinctions – all these are part and parcel of the ceaseless changes of language…they are only that part against which the best public sentiment, a healthy feeling for the conservation of linguistic integrity, arrays itself most strongly. (Whitney 1868:84-5, cited by Labov 2001:30)

Contrary to this simplistic view is that of those linguists who considered the and to highest group in the social hierarch as the starting point of linguistic change. Among them Labov mentions Tarde and Wundt, and the Project on Linguistic Change and Variation in Philadelphia (LCV), especially designed to find the social groups responsible for language innovation. Labov (Labov 1963, Labov 1966a, and Labov 1965) made the opposite prediction to the early 'top or bottom' theories of the causes of language change in his model of the mechanism of language change. As Labov stated:

> (…) change could originate in any social group, and following Sturtevant's suggestion, would spread gradually through each neighboring social group until it reached in one form or another all members of the community" (2001:31).

In fact, in his discussion of the social location of the social group responsible for innovations Labov located the innovative group in upper working class or lower middle class, which he considers the centrally located groups (Labov 2001:31-2). Resulting from this was the postulation of the so-called 'curvilinear hypothesis' which associates stable sociolinguistic variables with a monotonic social class distribution, while a monotonic distribution in age groups is related to a curvilinear pattern in the socioeconomic structure (Labov 2001:32).

Important, however, is to bear in mind the third postulate of language theory as cited by Chambers (2002:349):

> "Not all variability and heterogeneity in language structure involves change, but all change involves variability and heterogeneity" (Weinreich, Labov and Herzog 1968:188).

Therefore, argues Chambers (2002), language variation in correlation with social factors enables predictable patterns. Consequently, the implication of the postulate above according to Chambers is that language change is one kind of language variation determined by social forces. In association with social class and sex, age is considered one of the key social categories responsible for language change.

In his article on patterns of variation and change Chambers (2002) claimed that language variation may also indicate instability and change and, therefore, it is correlated to age. As he explains,

> … the change reveals itself prototypically in a pattern whereby some minor variant in the speech of the oldest generation occurs with greater frequency in the middle generation and with still greater frequency in the youngest generation. If the incoming variant truly represents a linguistic change, as opposed to an ephemeral innovation as for some slang expressions or age-graded change, it will be marked by increasing frequency down the age scale.(Chambers 2002:355)

As De Paiva, et.al (2003:13), citing Labov 1963) argue, according to the apparent-time hypothesis, language change can be understood in its course, that is, across different generations of speakers. This view is a clear shift from the historical perspective of language change theory based on comparison of different historical moments in the history of the language. In the conclusion of the article on "Real and Apparent Time" Guy Bailey (2002:329) recommends a caution regarding theories, given the fact neither offers a direct reflection of change in progress. In fact, the main assumption of the apparent time hypothesis is the argument of the stability of linguistic variable, that is, the absence of change. According to this perspective at a certain pointing after adulthood period dividual linguistic

practices remain stable (Labov 2001:85-6). This claim, argues Schilling-Estes (2002:309) "has yet to be tested in full".

3.6 An overview of models of language acquisition especially Chambers' (1998) eight principles of language acquisition

The interest on the subject of *languages in contact* has motivating many investigations, with particular reference to the pioneering study conducted by Uriel Weinreich in 1953. Indeed, it is a common sense that the processes taking place during *language contact* substantially are result of individual bilingualism, which is required in order to enable communication between two speakers of two mutually unintelligible languages. The fact that individual speakers can manipulate two or more varieties enables influence and interaction and the languages in contact may become changed linguistically.

As stated in chapter one of this dissertation, this study is also interested in the way in which contact may lead to variation and change. However, it deals not with mutually intelligible dialects, but with mutually unintelligible *languages in contact*. This situation of *language in contact* I will deal with is concerned to Mozambique and it is a result of European colonial expansion. The spread of colonial languages, in the case, the Portuguese, created a society where Portuguese, a high-prestige European language, coexists in a dominating position with the native languages spoken by the majority of the Mozambican, the conquered people. This section summarizes two of the eight principles of dialect acquisition tested in this study. Chambers (1998) postulated these principles, as a result of his developmental study carried out with six Canadian youngsters in southern England in 1983.

In Principle 4, Chambers points out that *Acquisition of complex rules and new phonemes splits the population into early acquirers and later acquirers*. According to this principle the acquisition of complex phonological features is age-graded and splits acquirers into early and later acquirers as age graded. Age becomes critical in dialect acquisition as it gives advantages to youth in both first-language and second-language acquisition (Chambers 1998, citing Long 1990). However, based on other studies of dialect acquisition (Sibata 1958 and David & Montgomery 1988), he also noted that, although there is a rough correlation between early acquires and later acquirers with younger and older subjects, it is difficult to

draw a clear age boundary between the two categories, the acquisition and age, making it prudent to resist identifying early acquirers by age group. Indeed, quoting David & Montgomery, Chambers (1998) admits a slight initial advantage of old children over younger ones when both are within the age limits of the "language-forming period". Moreover, Chambers concludes that notable sociolinguistic studies reveal that late learners may never master complex rules of new accent. In my study, this situation seems to be also extensive to early acquirers.

In Principle 6, Chambers claims that *Phonological innovations are actuated as pronunciation* variants. This principle is related to the theory of language change, also known as Lexical Diffusion (Wang & Cheng 1970, Chen & Wang 1975, Chambers & Trudgill 1980: 174-80, Ogura 1990: Ch. 6, all cited by Chambers 1998:160). According this theory, phonological innovations result from the acquisition of particular instances of the new rule or phoneme, word by word, and they only become rule-governed or systematic after the acquisition of the most important instances (Chambers 1998:164).

Trudgill (1986, quoted by Chambers 1998), postulated the speakers' efforts at accommodation as motivation for phonological innovations as follows:

> The point is that during accommodation speakers do not modify their phonological systems, as such…. Rather, they modify their pronunciations of *particular words*, in the first instance, with some words being affected before others. Speakers' motivation … is *phonetic* rather than phonological: their purpose is to make individual words sound the same as when they are pronounced by speakers of the target variety. Trudgill (1986:58)

There are two important aspects to mention here. First, this study, as stated before, deals with mutually unintelligible languages in contact. Second, the acquirers cannot benefit from the pronunciation of the speakers of the original target variety, as they are not available. Therefore it is safe to assume that the acquisition of the new dialect is fully determined by a multiplicity of other factors including that of the interference from the African languages also in contact. Next section is dedicated to the emergence of *new varieties of Portuguese*.

3.7 The spread of European languages and the emergency of new varieties

As the result of European colonial language spread some languages, because of their wide spread, became the important means of the world verbal communication for multiple purposes and functions among conquered non-native speakers. In these cases many language varieties, already modified from those of native speakers, are acquired or learned in different formal and informal contexts and are used as second or third languages (cf. Platt et al., 1984: 4). They may even, at a later stage, stabilise as first languages, both, by monolingual and by bilingual children.

Therefore, Ferguson (1983:viii) argues that from the perspective of general theory of language the native speaker or non-native kind of language use is equally crucial to the linguistics research. One of the reasons for this claim is that language variation between two languages is as important as it is in one single language to account for 'phenomena like language acquisition, language convergence or language shift' in relation to the learnability of natural languages and the nature of language change.

Considering different scenarios in which languages in contact can influence each other, many controversial claims are made and no consensus has been reached yet. One of the reasons why so little agreement has been reached regarding the linguistic consequences of language in contact derives from the divergent views about the nature of language and the type of language contact situation itself (Appel an Muysken 1987:153). One of the general claims that have been made is that of language convergence. This hypothesis suggests that in many cases, unrelated languages spoken in the same space for a long period of time have developed similarities at different levels of their structural systems. One of the most studied type of convergence is that of phonological convergence in which the sound systems of the languages in contact become more similar, without showing any clear influence between them (Appel and Muysken 1987:154). The example below in Table 24 illustrates roughly a situation of dialect variation in which the pronunciation of fricative palato-alveolar /ʒ/ is parallel in EP, MP, giTonga and xiChangana:

Languages	EP	MP	giTonga	xiChangana	English
South of Mozambique	laran[ʒ]a	laran[ʃʒ]a	lidalan[dz]a	laran[dʒ^{h}]a	'orange'

Table 24: Dialectical variation of fricative /ʒ/

There are many other situations documented in literature reporting grammatical futures that are shared by languages in their later stage of contact (Jackobson 1931, Birnbaum 1966, Civian 1965, all cited by Appel and Muysken 1987:155). Although there is no one single explanation about why different kind of convergences take place, in the above context of phonetic convergence it seems plausible to assume that these similarities in the pronunciation of fricative palato-alveolar /ʒ/ emerged as a result of language contact.

Another claim made about linguistic consequences of language contact is related to the so-called *substrate influence* (Appel and Muysken 1987:157). According to these authors when a foreign language is spoken in a new region different from that of its origin and, because of its prestige, the new speakers adopt it as their second language, and then the original language or languages by these speakers in many ways can influence this language. Additional it is also expected the inverse situation in which the new language influences the native language or languages of the speakers.

In fact, especially in Africa, in many of these multilingual contact situations the interaction between these varieties or cluster of varieties of a language, also called extraterritorial languages (See Lass, R in Hickey, R 2004:365), and the autochthonous languages result in sociolinguistic phenomenon showing a continuing process of mutual influence. Bokamba (1983:77) notes that for Africa such processes are not yet well studied if not simply neglected.

3.8 Second language acquisition and learning process: implications for language variation

The process of second language acquisition is often associated with many different bilingual experiences, where the general purpose is simultaneously to ensure the maintenance of the native language and/or to continue to develop the competence in native language by

the learners while acquiring a second language (Larsen-Freeman and Long 1991:1). The immersion programmes are also another form of second language acquisition most popular in Canada and the United States. In these programmes English-speaking children receive their initial instruction in a second language, after which the content courses are taught in the native language. Another form of second language acquisition, but different from the previous experiences, is concerned with the well-known acquisition of the "other tongue" (Kachru, 1983), with reference to the imposed ex-colonial languages often used as superior and powerful for religious purposes, cultural subjugation and colonization.

Although some of the new nations, after the colonial era, have adopted an indigenous language as the medium of instruction in primary schools, most of education both in higher schools and tertiary institutions remain through the medium of European languages. Given the growth of educational access for the majority of local populations, many children are now acquiring the ex-colonial languages either as their first or as their second languages (cf. Platt *at.al* 1984). In many cases some the resulting situation of differentiated use of languages is a situation of *diglossia* (See Ferguson 1971), where speakers use their languages and the *other* tongues in different non-overlapping contexts.

For some contexts, however, there is a clrear evidence of language shift, with younger people gradually replacing their indigenous languages by the ex-colonial language associated with social mobility and economic opportunities by the parnts. In a later stage, a result of language contact can be totally a new language in a form of a pidgin. This is a situation in which a reduced linguistic system used occasionally among speakers of different languages and that it is not also a mother tongue. Another rsulting linguistic form is a creole, a later stage of a pidgin and that had acquire native speakers already (Appel and Muysken (1987:175). Besides the parable of disaster of Bickerton (1975), cited by Appel & Muysken (1987:175), or a migration of socially dominated groups, I believe that in a very post-colonial context other sociolinguistic factors can motivate the emergence of pidgin and creole languages. This can be, for example, in some contexts of interaction among children often from environments where literacy is rare. These children because of the language attitudes of their parents do not speak properly any of the languages they are trying to employ in their communication, namely their own mother tongue and the ex-colonial prestigious language.

Interesting is the fact that sometimes even children from educated and/or European descendant families involved in such interactions frequently use newly coined words and other new grammatical structures. Usually these innovated forms are taken from the mother tongues of the native children. However, whether there is pidginization and/or creolization still needs to be studied carefully.

As in many diglossic speech communities, in Mozambique many distinct but related African languages and language varieties are used parallel with Portuguese, each with a clearly defined role. In this context of diglossia, Portuguese, as the so-called superposed variety, is assigned the H ('high'), function, and all African languages are assigned the status of L ('low') varieties. Thus, many people in Mozambique speak their local mother tongue at home, among family or friends of the same variety area, whereas Portuguese is used more often to communicate with speakers of other varieties or in official and administrative domains. There is, however, a great overlapping of the two sets especially on family occasions, thus rendering the diglossia model less applicable. At the official level the use of H is so strongly associated with high prestige that many Mozambicans often say they don't speak any 'language', except the 'dialect'. By this they mean that they don't speak Portuguese, but are in highly fluent in their local Africa mother tongues. This attitude is not only associated with uneducated speakers. It is a result of the stigmas of the colonial past, where Portuguese was the only language considered as such, and therefore somehow more superior and the language of culture and for social mobility.

This differential status carries implications for the method of acquisition of the languages at present. In effect, children learn their mother tongues, naturally by their exposure at home without any explicit explanation of grammatical structures. Platt *at al.* (1984:3) referring to diversity of ways many ex-colonial languages were and are acquired in the independent nations today with the increasing of education facilities, argue that, in addition to the fact that many second language teachers have not more native-like command, the school is also not the only way through which the children acquire the second language: They heard it from older children in the playground, on the way to school and on the way home. In fact, many children learn the second language at home from their older brothers and sisters (Platt *et.al* 1984:4). This situation in many cases led to the emergency 'New Varieties' of the metropolitan languages.

CHAPTER 4
METHODOLOGY AND DATA COLLECTION

As I stated in chapter 1, this study investigates phonological variation in one of African variety of Portuguese: the Mozambican Portuguese. The main source of data of this study is the linguistic database known as *Panorama do Português Oral do Maputo (PPOM)* (Stroud and Gonçalves 1997), a linguistic survey gathered in the context of the study of oral Portuguese in Maputo.

The outline of this chapter is as follows: In section 1 I present a brief characterization of the design of this study, the factor groups and the selection of populations and sampling. In section 2 I describe the information sampling. In section 3 I deal with the transcription and quantification of linguistic data, the procedures and constraints. Finally, in section 4 I summarize the quantitative model based on VARBRUL (Cedergren, H.; and Sankoff, D. (1974); Pintzuk 1988).

4.1 Design of the study: Selection of population and sampling

For the purpose of this study, I explore methods of investigation at the micro-variation level (Romaine 1982:66). The main objective is to determine in what way linguistic variables, i.e., the way in which sepeakers vary in their linguistic behaviour (Labov 1963 cited by Romaine 1982: 66, Labov 1980:2, and Chambers et al. 1980:49), are affected by extralinguistic factors, handling quantitative data without overlooking qualitative tendencies in them. Studies on language variation have demonstrated that linguistic variables correlate very closely with a number of extralinguistic factors. For instance, socio-phonetic variables /ʀ/ and /ʎ/ in MP may be conditioned both by the age and sex, on one side, and, on the other, the method of instruction and the local of residence of the speaker. It is not my focus in this study the influence of internal factors.

The geographical base of this research is comprised of five residential quarters of the city of Maputo, the capital but also the biggest urban centre of Mozambique. Three of them are considered suburban areas and the remaining two urban (Stroud and Gonçalves 1997:56,

quoting *Relatório* 1992). Of approximately 1,099,102 inhabitants, the city of Maputo has been able to maintain the highest level of services and formal and informal commercial activities, making it the most attractive urban immigration centre of the country. One of the linguistic implications of this situation is that Maputo is today also the most multilingual conglomeration in Mozambique; where Portuguese and Bantu languages co-exist as illustrated in chapter 2, section 3, previously. The selection of the city of Maputo was determined by two key factors, namely: (i) that it is the major and more diversified community of speakers of Portuguese in Mozambique, and (ii) that Maputo is the major space where Portuguese is the predominant language of communication (IBRAIMO 1994:33, cited by Stroud & Gonçalves 1997:55).

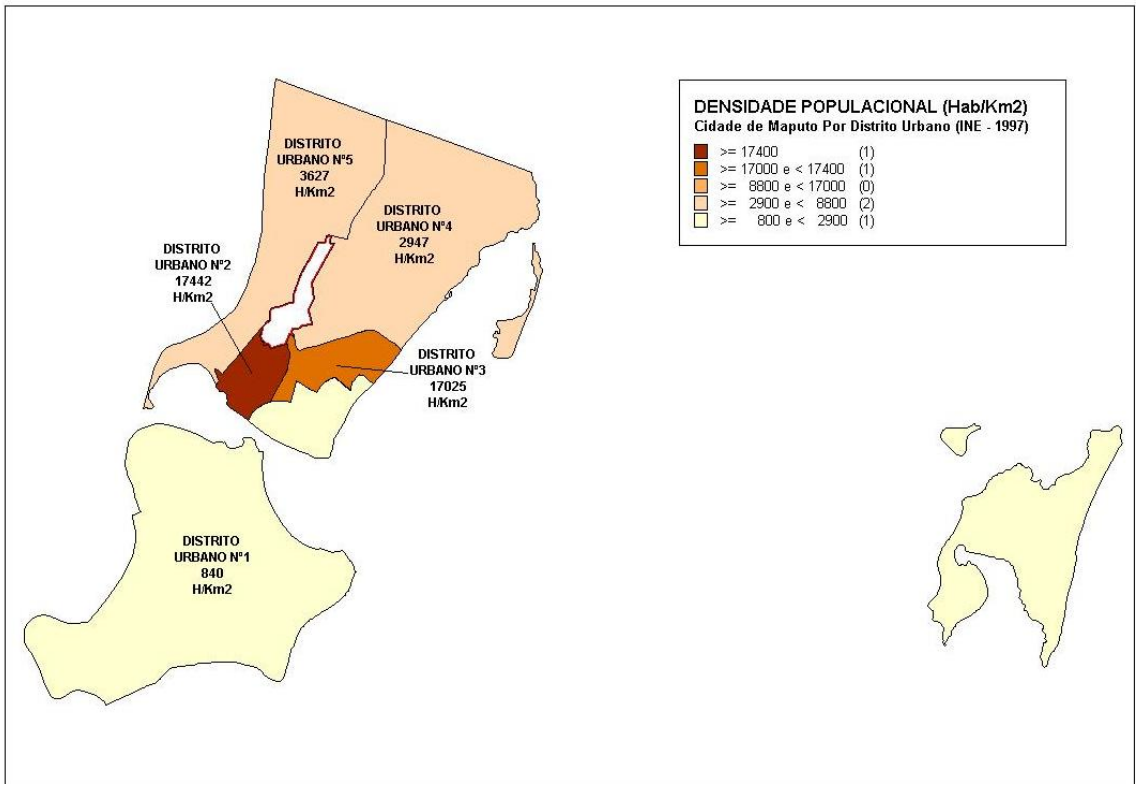
From the 1997 population census, to the last one in 2007 there is a significant increase in terms of the number of urban areas and in terms of the population. In effect, the population of city of Maputo increased from 966.837 inhabitants in 1997 to 1.099.102 in 2007. The number of urban areas icresed also from a total of five in 1997 to seven in 2007. According to de Araújo in his comparative study of the population of Maputo and Matola cities (de Araújo [n.d.]), despite this multiplicity of geographic spaces, there is no discontinuity between them. He argues that besides their geographic contiguity and the physical continuity, Maputo and Matola have intense daily social and economic interrelations, which make them a whole unity. However, although these claims are true from a very restricted geo-spatial and economic point of view, from a linguistic viewpoint these urban spaces have different and diverse dynamics.

According to Stroud and Gonçalves (1997:56), different residential quarters of the city of Maputo can be divided into two groups, namely: (i) residential urban quarters and suburban residential quarters. The later comprises *Maxaquene, Chamanculo* and *Mafalala* and the former includes *Alto-Maé* and *Polana Cimento*.

For the present research, I selected five residential quarters based on two main reasons. According to the authors of the linguistic data used in this study, the first reason is that the inhabitants of *Maxaquene, Chamanculo* and *Mafalala* residential quarters are more stable and relatively immune to the influence of population exodus from rural to urban areas, caused by civil war and drought.

Furthermore, the authors claim that inhabitants of these residential quarters are also: (a) L1 speakers of a local Bantu language of Maputo, namely xiRhonga or xiChangana (*Chamanculo* in *DU2* and *Maxaquene* located in *DU3*) or (b) L1 speakers of a non-local Bantu language of Maputo, in the case speakers of eMakhuwa, spoken in the north part of Mozambique (like in *Mafalala* residential quarter). This fact determined the selection of the second group of residential quarters, namely, *Alto-Maé*, a residential quarter at the border between urbane and suburban residential areas, and Polana Cimento, a typical urban residential quarter of the city of Maputo. These two residential areas are located in *DU1*. Also, according to Stroud and Gonçalves (1997:57), *Alto-Maé and Polana Cimento* differ from suburban areas because of the intense use of Portuguese by the residents in their daily verbal interaction.

Therefore, the choice of different contact areas between Portuguese and local Bantu languages, on one side, and the selection of areas of differentiated usage of Portuguese in daily verbal interactions on the other side, can bring to light important aspects of the nature of socio-phonetic and phonologic variation of Portuguese in city of Maputo and the whole country. The population density of the city of Maputo is as shown in Map 6. Many authors consider the DU1 as the most urbanized area of Maputo. However, as the map below illustrates, this administrative area is in contact with *DU2* and *DU3*, considered suburban areas.



Map 6: Population density (Inhab/Km^2) of Maputo by urban districts (*INE*, 1997)

(In http://www.iid.org.mz/iucn/html/cidade_de_maputo.html**)**

Although I partially agree with Stroud and Gonçalves (1997:57) when they suggest that urban areas differ from suburban ones because of their intense use of Portuguese by their residents in their daily verbal interaction, I consider that it is important to distinguish between "frequent" and "intense" uses. In effect, *frequent* use of a language, on one side, means its regular use, in other words, the number of times the language is used in a particular period of time, and this seems to me the case of the use of Portuguese by the inhabitants of *Alto-Maé* and *Polana Cimento*. The frequency can be deduced from the number of the speakers using relatively the same language: the bigger the number of L1 speakers, the higher the frequency.

On the other side, *intense* use of a language means a very great or extreme use of a language, and this is the situation of suburban residential quarters of *Chamanculo, Maxaquene* and *Mafalala*. The intensity is measured by the number of the speakers of different languages trying to communicate in a common language: the bigger the number of the users of this lingua franca, the higher the intensity.

As illustrated in the map 6 above, according to the 1997 population census, the three selected suburban residential quarters for this study, namely *Chamanculo* in *DU2* and *Maxaquene* and *Mafalala* in *DU3*, are situated in the areas with the highest density of the city of Maputo. Coincidently, these areas are also neither the most populated nor geographically the biggest areas of the city. Note that this situation changed drastically in the last ten years. In effect, after the last population census in 2007, these districts became part of the most populated districts of Maputo, as indicated in Table 25 below.

	TOTAL	**%**	**MALES**	**%**	**FEMALES**	**%**
TOTAL	**1.099.102**	**100,0**	**531.794**	**48,4**	**567.308**	**51,6**
DISTRITO URBANO No. 1	106.346	100,0	51.515	48,4	54.831	51,6
DISTRITO URBANO No. 2	155.462	100,0	76.199	49,0	79.263	51,0
DISTRITO URBANO No. 3	223.688	100,0	109.2	48,8	114.488	51,2
DISTRITO URBANO No. 4	293.768	100,0	141.302	48,1	152.466	51,9
DISTRITO URBANO No. 5	293.998	100,0	141.166	48,0	152.832	52,0
DISTRITO URBANO No. 6	20.629	100,0	9.964	48,3	10.665	51,7
DISTRITO URBANO No. 7	5.211	100,0	2.448	47,0	2.763	53,0

Table 25: Population of the city of Maputo by *DU*

Consequently, all these districts constitute the pivot of informal commercial transactions of the city and, therefore, are daily crossed by millions of sellers and buyers speaking different languages, and who find the Portuguese the only expeditious means of communication.

The two selected urban residential quarters, respectively *Alto-Maé* and *Polana Caniço*, were situated in the most populated area of the city, the "*Distrito Urbano N^{o} 1*", but simultaneously the least dense place of Maputo. After the population census of 2007, this district is one of the less populated areas of the city. According to Stroud and Gonçalves (1997:55), given the fact that the majority of the informants spoke Portuguese as their second language, it was a necessary condition that they should have achieved at least an equivalent to standard 4 of more higher of formal education, in order to enable them to communicate satisfactorily in Portuguese about daily activities.

The empirical data of this study was drawn from the linguistic analysis of 11 hours of recorded interviews from approximately 130 hours of recorded group discussions and recorded interviews. The selected 20 speakers from a total of 100 were all resident in Maputo City. The selection of the sample comprised male and females within the age bracket 16 to more than 55 years. Note that the oldest speaker selected to this study is 57 years of age.

One important factor explored by Labov (1966, 1972a), Trudgill (1974b) and G. Sankoff (1974) is the socio-economic category of the subject population. However, the authors of the PPOM project avoided in their research categorizing sample population into such pre-established socio-economic categories. In many contexts, one of the problems related to the use of this factor is the difficult of clearly satisfying the Mozambican society on the base of a combination of both economic and cultural background. Table 26 below summarises the sociolinguistic information about twenty one participants in this study. The source of this information is the *PPOM* linguistic database with personal changes.

Informant's Code	Fictional Name	Sex	Age	Education[1]	Mother tongue	L2	Other spoken or known languages
MF12RAF	Rafaela	F	16	3ª classe	Makhuwaa	Port.	
CH9ADA	Ada	F	22	7ª classe	Rhonga	Port.	Chona
MX9KUT	Alex	M	29	10ª classe	Rhonga	Port.	
MF11VIR	Vera	F	46	4ª classe	Makhuwa	Port.	Rhonga
CH18SAC	Sara	F	28	9ª classe	Rhonga	Port.	
AM 22PEI	Pedro	M	26	3º Ensino Técnico	Sena	Port.	Ndau, Chona
MX5HOR	Horácio	M	32	3ºInstitituto Industrial	Changana	Port.	English
MX23NAU	Nora	F	46	3ª elementar	Changana	Port.	Copi
PC8COS	Costa	M	55	11ª classe	Rhonga	Port.	
CH7OJM	Joel	M	33	11ª classe	Rhonga	Port.	French
AM8JOA	João	M	46	4ª classe	Tonga	Port.	Tshwa, Rhonga, Changana, English
CH16FEC	Félix	M	45	6ªclasse	Rhonga	Port.	
CH2NEL	Nelson	M	23	11ª classe	Rhonga	Port.	Changana
CH10LUM	Lúcio	M	25	7ª classe	Rhonga	Port.	German
MX2SIM	Simeão	M	22	7ª classe	Changana	Port.	
MF27IMO	Moisés	M	56	4ª classe	Rhonga	Port.	Changana
MF25ABU	Abú	M	43	2º ano comercial	Rhonga	Port.	
MF5ZAF	Safira	F	56	4ª classe	Makhuwa	Port.	Changana, English
AM28VER	Virgínia	F	57	8ª classe	Changana	Port.	Rhonga Tonga
MX3MAR	Maria	F	29	6ª classe	Rhonga	Port.	German
AM21PAU	Paulo	M	26	9ª classe	Tonga	Port.	Changana, Rhonga, Copi

Table 26: Sociolinguistic data of informants

[1] The standards are expressed according to the colonial education system and the New National Education System *(Sistema Nacional de Educação – SNE).*

4.2 The phonological variables

This section deals with the way the variants in the data base were classified in preparation for later statistical analysis via VARBRUL. One of the primary aspects to deal with in sociolinguistic research is the selection and delimitation of the linguistic variable and its variants as the basic working tool (Hazen 1998:221), and Appel and Muysken 1987:143). Hazen, citing Wolfram (1993:195-200), defines a linguistic variable as a group of variants operating in a linguistically defined set, arguing then that it is crucial to explore its linguistic and sociolinguistic aspects.

From the point of view of sociolinguistic selection of the linguistic variable the linguistic variants are organized according to their social relevance in the community of speakers (Hazen 1998:221). For example, besides their linguistic relations the two forms of the phoneme /ʀ/ in MP have social importance. Although in this study I am not arguing specifically about the velarization of /ʀ/ it is important to note that in fact in the MP variety the speakers tend to be socially stratified according to this process indicating clearly that the velarization of /ʀ/ is a sociolinguistic marker. Gumperz (1958), cited by Wardhaugh (1992:143-144), in his study of linguistic usage in Khalapur village in the north part of Delhi in India demonstrated how linguistic variation is related to social variation.

Chambers & Trudgill (1980:60), cited by Walters (1988:126), in their turn define a linguistic variable as a linguistic unit with two or more variants, that is, different realisations like in words with more than one pronunciation. These linguistic variables are in a covariation with other social and/or linguistic variables. According to them (citing Labov 1972b:8 and Labov 1966:49), linguistics variables preferably should be frequent in their occurrence, socially stratified, structured units and quantifiable on a linear scale.

Considering the correct analysis of these linguistic units as the most crucial moment in sociolinguistic research, Labov (1972b:72), cited by Walters (1988:126-127), postulated three important steps in the definition of linguistic variable, namely: (a) the enumeration of the number of contexts in which the variable occurs, (b) the identification of many phonetic variants as much as possible, and (c) the need to assign the variables with quantitative indexes. However, Hudson (1980:145-6) identified the following three constraints regarding

Labov's steps. The main difficult task is to define the environments, to decide the forms to be counted as representing variables as well as the distinction of a 'reasonable, accurate, and appropriate number of variables'. Other problems pointed by Walters (1988:123) are related to the difficult to select and analysing linguistic variables and the limitations of the applicability of notion of linguistic variable in other aspects of linguistics rather than these related to phonology.

4.3 Criteria for selection

The two selected socio-phonetic variables are /ʀ/ and /ʎ/. The candidate variants for /ʀ/ are the following. First, is the uvular fricative [ʁ], with a prestige feature and associated with younger and (some) educated people. The tap or flap flap [ɾ], alternatively also trilled, is another selected variant. The third variant is the trilled [r], frequently also flapped. The last two variants, namely the flap trilled or the trill flapped have both low prestige and, therefore, associated with low educated people and/or Bantu mother tongue speakers. The fourth selected variable is the lateral palatal /ʎ/, which variants in some south MP varieties are the lateral alveolar [l] and the lateral fricative [ɬ], both with low prestige and associated with Bantu mother tongue speakers.

Because of the quantity of time and effort involved in identification of phonological variables, I decided to limit the number of tokens to a maximum of 100 where there were more tokens than this number.

One of the case of this study is the pronunciation of /ʀ/ in MP at the initial and middle position of the word. The following tables 27, 28 and 29 represent the resulting situations and respective linguistic variants.

(i) Initial and middle flap realizations of uvular fricative [ʁ] or alveolar trill [r]

The initial and middle flap [ɾ][1] realizations of uvular or trill is summarized in Case One and Case Two of this study, respectively [r][2] or [ʁ], and shown in Table 27.

Case One and Case Two – Initial and Middle flap realization of Uvular or Trill	
Tap or flap [ɾ]	1
Uvular fricative [ʁ] or Alveolar trill [r]	0

Table 27: Flap realization [ɾ] of uvular [ʁ] or trill [r]

(ii) Middle uvula [ʁ]or trill [r] of flap [ɾ]

Table 28 indicates Case Three of middle uvular [ʁ] or tril [r] realization of flap [ɾ].

Case Three – Middle uvular or trill realization of flap	
Uvular fricative [ʁ] or Alveolar trill [r]	1
Tap or flap [ɾ]	0

Table 28: Middle uvular [ʁ] or trill [r] realization of flap [ɾ]

(iii) Middle lateralization or fricatization of the phoneme /ʎ/ in some MP varieties

[1] Tap or flap [ɾ] in normal Potuguese of Portugal and Brasil (Martins, 1988:171).

[2] Alveolar trill [r] in normal Portuguese of Portugal, Rio Grande do Sul and other regions of Brasil, also articulated as uvular [ʁ] in Lisbon, Rio and in wide areas of Brasil (Martins, Maria Raquel Delgado, 1988:171).

Case Four of middle lateralization [l] or fricatization [ɬ] of lateral palatal /ʎ/[1] is shown in Table 29 below.

Case Four – Middle lateral or fricative realization of palatal	
Lateral alveolar [l] or fricative [ɬ]	1
Lateral palatal [ʎ]	0

Table 29: Lateralization [l] or fricatization [ɬ] of palatal [ʎ]

Note that there is no particular significance to the 1 or 0 in the above labeling. However, I tried to give an "application value" of 1 to what seemed the more common realization and 0 others.

4.4 Sociolinguistic Factors

In order to cope to the social relevance of the linguistic variable, the participants were selected according to some social characteristics such as place of residence, education, gender and age as summarised below in Table 30.

	Residence	Education	Sex (Gender)	Age
Factors	S – Suburban area	P – Primary Level	F – Female	y – younger
	U – Urban area	H- Higher Level	M – Male	m – middle-aged
		T – Tertiary Level		o – older

Table 30: Sociolinguistic factors

Additional were considered other social factors such as the L1 of the participants as well as the distributional context of the analysed variables, namely the initial and middle position within the word. These factors were not included in the data run by VARBRUL given the fact that hey are transversal (all participants are L1/Bantu language speakers) or were analysed separately in case the distributional context.

[1] Normal Portuguese. of Portugal and Brazil (Martins, 1988:171)

4.5 Information sampling

The sampling procedure was basically to select randomly twenty informants since each participant in the *PPOM* linguistic survey had a probability of being selected. In fact, according to Stroud and Gonçalves (1997:56-57), the target population is distributed within different age brackets and residential areas. The majority of the population speaks a Bantu language as L1 (92%) and had no tertiary education (94%). More than a half of a total of 100 males and females (62%) access to the Portuguese language via school and lives in Maputo for more than ten years (77%).

For this study I selected eight females and 12 males. The discrepancy in terms of the number of participants by sex is related to the lack of recording quality of some of the tapes. In fact, initially was selected equal number of participants for each sex. However I believe that this situation will not affect substantially the final findings given the fact that sex seems to not be the most prominent social factor in the variation of MP. The main kind of information comprises responses to interviews and group discussions.

The interviews consisted of the informant's recorded responses to a guided questionnaire about five pre-selected general and accessible themes for all ages, sex, social origin and education. These interviews were carried out by fieldworkers totally unknown by the informants. Additionally, in most of situations the fieldworkers had to overcome a considerable number of constraints about where to record the interviews. Therefore it was difficult to obtain the very high quality tapes necessary for the detailed phonological analysis intended, due to surrounding noisy and technical experience by the fieldworkers. Besides, it is also important to notice that informants of different background (e.g., education, age, sex) responded differently to the linguistic stimulus. Table 31 shows the sample distribution according to different social factors.

	Alto-Maé **(A)**M			**M**afalala M(**F**)			**M**axaquene M(**X**)			**Chamanculo** (**C**)H			**P. Caniço** (**P**)**C**		
	(**Y**)oung	(**M**)iddle	(**O**)ld	(**Y**)oung	(**M**)iddle	(**O**)ld	(**Y**)oung	(**M**)iddle	(**O**)ld	(**Y**)oung	(**M**)iddle	(**O**)ld	(**Y**)oung	(**M**)iddle	(**O**)ld
	16-35	36-55	56 +	16-35	36-55	56 +	16-35	36-55	56 +	16-35	36-55	56 +	16-35	36-55	56 +
Urban	Pedro (M) Paulo (M)	João (M)	Virgínia (F)											Costa (M)	
Suburban				Rafaela (F) Ada (F)	Vera (F) Abú (M)	Moisés (M) Safira (F)	Alex (M) Horácio (M) Simeão (M) Maria (F)	Nora (F)		Sara (F) Joel (M) Nelson (M) Lúcio (M)	Félix (M)				

Table 31: Sample Distribution by place of residence, age brackets and sex

4.6 The quantitative model based on VARBRUL

According to McEntegart, Damian and Le Page, R. B. (1982:109) one of the common difficult related to all quantitative linguistics is the *identity* of the linguistic units to be counted, given the fact that what seems to be systematic in idealized linguistic studies differ considerably from that of real life and in a contact situations. In fact, they argue that all linguistic units (e.g., phonemes, morphemes and lexemes) maintain their identities while the system is being manipulated for descriptive purposes, partly because the speech and semantic functions are defined by the context of use. Although Labov and Trudgill highly recommended the work with phonological variables, it requires both "very good quality recording and a very careful, tedious transcription" (McEntegart, D. and Le Page, R. B. 1982:110).

The increasing in the use of statistical tools in the variationist studies especially in the paradigm of urban sociolinguistics' research has raised many difficulties for linguists because of their lack of training in mathematics (Sankoff 1978:xiii, cited by Walters 1988:128). Since its first use by Labov (1969), many efforts have been made in order to improve the variable rule analysis or VARBRUL (Cedergren and David Sankoff 1974, Kay 1978, Rousseau and D. Sankoff 1978a and 1978b, Kay and McDaniel 1979, D. Sankoff and Labov 1979 and D. Sankoff and Labov 1980b), all cited by Wlters (1988:128). VARBRUL is a statistical model which allows the measuring of effect on variation of different internal and external variables, comparing for example, situations where speakers alternate the uses of different linguistic forms with or not the same meaning.

The choice of either the one or the other form is conditioned by a variety of social factors traditionally believed as sufficient for an account for sociolinguistic variation (Fought 1999:5). In order to run the data in accordance with the VARBRUL methodology an application for multivariate analysis was developed, the GOLDVARB 2001, based on GoldVarb 2.0 program (Rand & Sankoff 1990, cited by Robinson, John; Lawrence, Helen & Tgliamonte, Sali 2001:4). This quantitative analysis model requires data organized in individual instances of a linguistic variable, or tokens, stored within token files (Robinson, John; Lawrence, Helen & Tgliamonte, Sali 2001:5). After checking the encodings for the factors VARBRUL indicates those with significant value to influence the variation. Knockout values (0% or 100%), i.e., categorical findings where all the tokens are accounted for by one

(Robinson, John; Lawrence, Helen & Tgliamonte, Sali 2001:24) are ‘knocked out’ (Kirk 1994:71) as they mean that there is no variation, the main aim of this model.

According to Kirk (1994:72), in order to calculate the significant factors VARBRUL deals with probabilities, in other words, the mathematical measurement of how likely a certain factor will occur, and principles of ‘maximum likelihood’, that is, the appropriateness of the analysis regarding the data (Robinson, John; Lawrence, Helen & Tgliamonte, Sali 2001:27). The multi-dimensional and regression nature of VARBRUL analysis allows to take into account all factors and to deal with them individually. The cross-tabulation alternative (Robinson, John; Lawrence, Helen & Tgliamonte, Sali 2001:21) is useful to determine significant causes of variation by intersecting factor groups each other. Table 32 shows the way the data were codified in order to be analysed by VARBRUL.

	Application values 1 (yes) / 0 (no)	Place of residence	Education	Sex	Age	Mother tongue	Structure characteristic
	Flap Alveolarization/Uvularization Lateralization/Fricatization	**A**lto-Maé **U**rban Mafalala **C**hamanculo **P**olana-Caniço **S**uburban Maxaquene	**P** – Primary **H** – Higher **T** – Tertiary	**M** – Male **F** – Female	**Y** – Young **M** – Middle **O** – Old	**L** – Portuguese L2 **P** – Portuguese L1	**I** – Initial **M** – Middle **F** – Final
Example: Rafaela String Code (1SPFYLI	**(1**	**S**	**P**	**F**	**Y**	**L**	**I**
Final string code	**(1 S P M Y** **0 U H F M** **T O**						

Table 32: Codification of data of the informant

In the Table 32 above are also specified the factor group 1 encoding the application values 1 and 0, where 1 is the dependent variable, i.e, the group which encodes the linguistic variable under investigation. Hence, in the token files, factor 1 means that the occurrence of the linguistic variable under investigation. In the example above it should mean that subject pronounces the /ʀ/ as a flap in the initial position of the word, while 0 means that the subject doesn't pronounce the /ʀ/ as a flap in the same initial position.

Although great achievements using the variable rule theory have been occurred, Walters (1988:131-2) notes that there are still problems with this methodology of language variation. A strongly critical view on the variable rule methodology comes from Sterelny cited by Walters (1988:133) as stated below:

> In my view, Labov has founded an industry rather than proposed a theory of a theoretical programme. For though there is an enormous amount of research along the lines he pioneered, no reason has been given for supposing that the research is more than the collection and display of data. (1983:66)

Few sociolinguists would accept this view. In this study I show that quantitative analysis can illuminate social variation and change in a way that other approaches in phonology do not.

CHAPTER 5
RESULTS AND DISCUSSION

In this study I have analysed the interaction of different forms of linguistic variable with particular external factors, i.e., the social phenomena that co-occur with a given variable form in order to understand how the linguistic variation is structured in the community of Portuguese speakers of Maputo City. Speakers in Maputo City, but not only, in many parts of Mozambique, indistinctly pronounce: (a) the uvular or trill /ʀ/ as flap [ɾ] at the beginning or at the middle of the word, (b) pronounce [ɾ] as uvular or trill at the middle position, and (c) lateralise [l] or fricatize [ɬ] the palatal /ʎ/ at the middle of the word.

The work to follow analyses four cases of co-occurrence of linguistic variables with four selected social factors, namely: residence, education, gender and age. For this purpose I will explore multiple regression analysis. My factor groups included for the present analysis in the best stepping-up and the best stepping-down runs are the same and their statistical significance is for all of them above the $p < 0.05$ level, the desired probability level (Bayley 2002: 127). The excluded factor groups are not significant to the variation under discussion.

For the interpretation of the results, among other aspects, I will consider the weights, that is, the numerical measure values between 0 and 1.00. These values indicate the strength or influence of each factor, in relation to other factors in the same group, on the occurrence of the linguistic dependent variable. Another aspect to consider is the input probability, which is the overall likelihood that speakers will chose the variant selected as the applicant value (say 1 in my study), i.e., the value that counts as an application of the "rule" under investigation (Bayley 2002: 127).

Case One

1 = Initial tap or flap [ɾ] or 0 = Initial uvular [ʁ] or trill [r]

Case Two

1 = Middle Flap [ɾ] or 0 = Middle uvular [ʁ] or trill [ʀ]

Case Three

1 = Middle uvular [ʁ] or trill [r] or 0 = Middle tap or flap [ɾ]

Case Four

1 = Middle lateral [l] or fricative [ɬ] or 0 = Middle palatal [ʎ]

My approach is based on the quantitative paradigm (Labov 1966, 1969a, 1972a, 1972b, cited by Bayley 2002:117). Therefore, the corpus of this study was transcribed and coded for VARBRUL, a method of multivariate analysis in quantitative sociolinguistics. The results are presented as follows.

5.1 Case One: Initial flap realization

1 = Initial flap [ɾ] and 0 = Initial uvular [ʁ] or trill [r]

The first variable to be discussed is initial flap realisation of the phoneme /ʀ/: 1 = [ɾ] or 0 = [ʁ] or [r]. My first computer run of the GoldVarb 2001 for windows included nine factors and ten cells of the first token file. This token file examines the likelihood of occurrence of initial flap realization in suburban and in urban areas of Maputo City by speakers of Portuguese. From 505-recorded tokens, 359 of them (71% of total) represent initial flap realization: 265 in suburban areas and 94 in urban areas, as indicated in Figure 1.

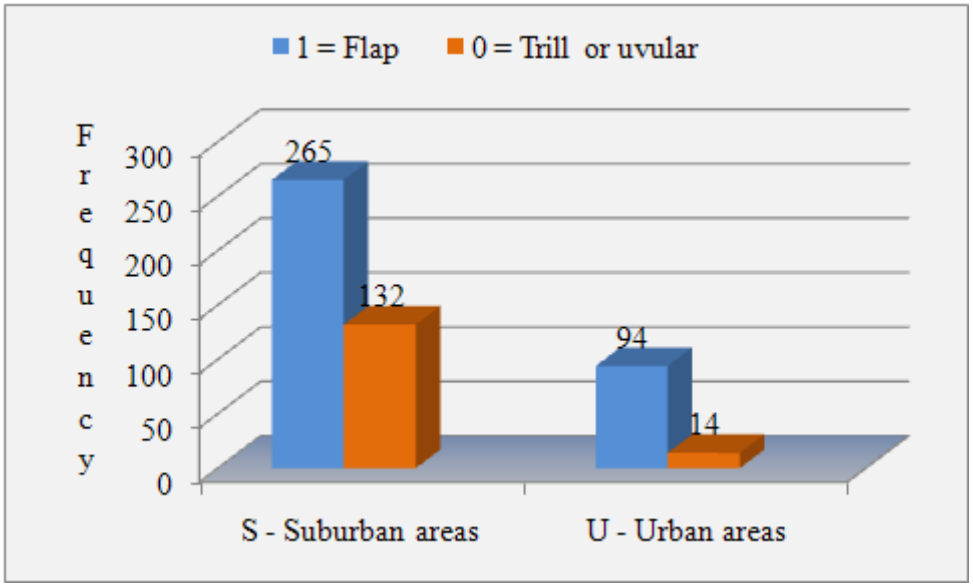


Figure 1: Realization of initial flap [ɾ] = 1 by residence

In terms of percentages as can be seen in Figure 2, in suburban areas speakers' realization of initial flap is 66%, while in urban areas speakers' percentage of initial flap realization is 87%.

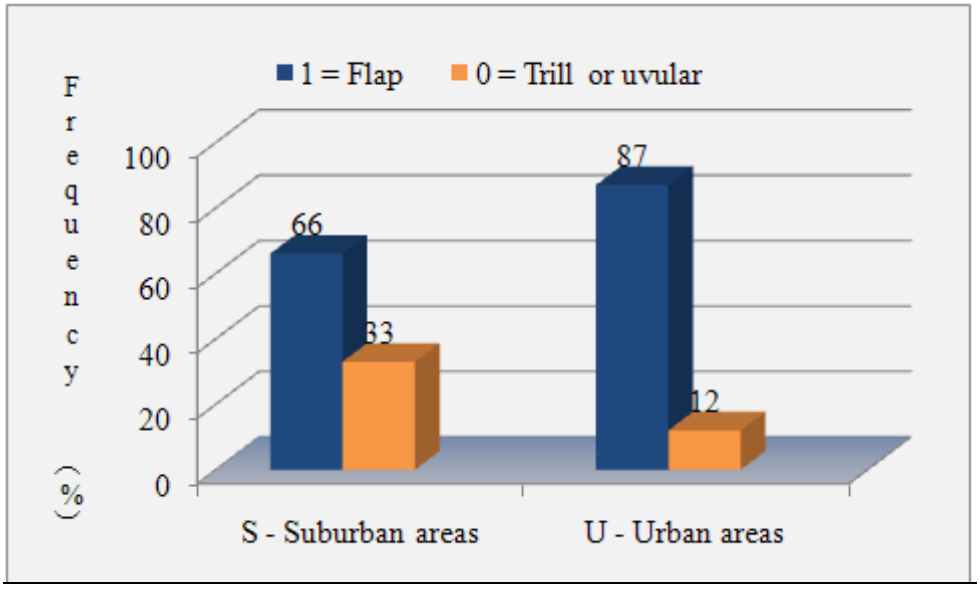


Figure 2: Percentages of initial flap [ɾ] = 1 by residence

Considering the number of tokens recorded in these two areas of Maputo City, urban areas show higher percentage of initial flap realization (87%) than suburban areas (66%)

The frequencies of the co-occurrence of realization of initial flap with speakers' level of education in Figure 3 below reveal that 222 tokens of 359 initial flaps are associated with lower educated people, while 137 tokens are related to higher educated speakers.

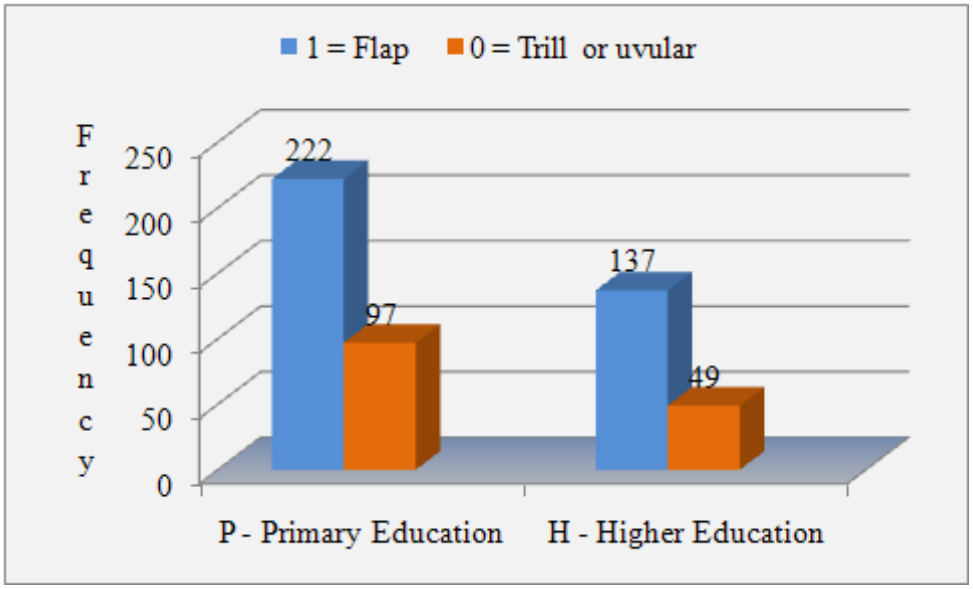


Figure 3: Frequency of initial flap [ɾ] = 1 by education

The comparative percentage of initial flap realization by speakers with primary level of education and by speakers with high level of education is indicated in Figure 4.

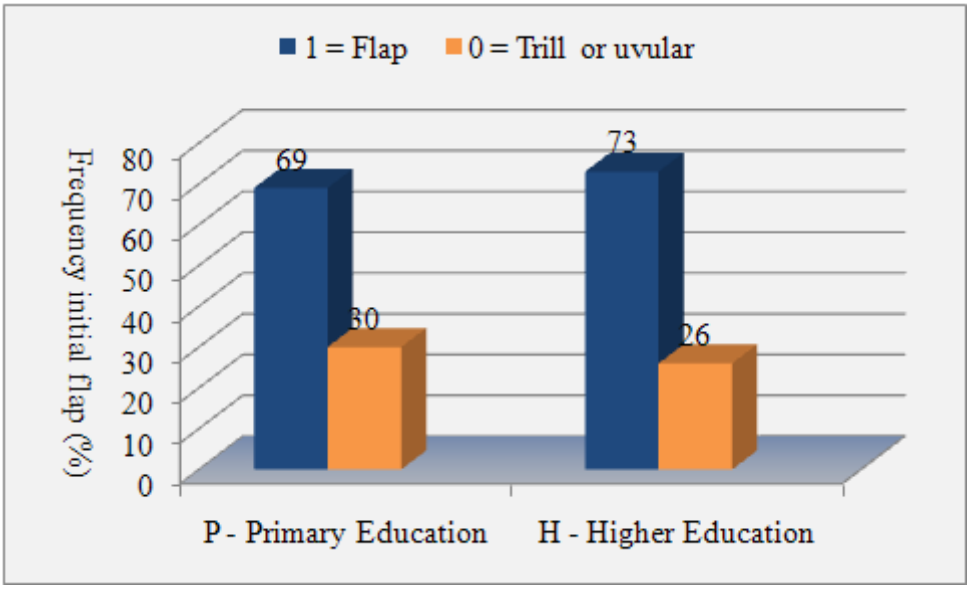


Figure 4: Percentage of initial flap realization [ɾ] = 1 by education

From 186 tokens recorded among speakers with high educational level 73%, reveals initial flap realization. Among speakers with lower level of education, from 319 tokens 69% represent initial flap realization.

Speakers of different sex relate another correlation examined here to the initial realization of flap. The results after the first computer run in Figure 5 of frequencies indicate that from a total of tokens of initial flap realization, 186 correspond to females and 173 to males.

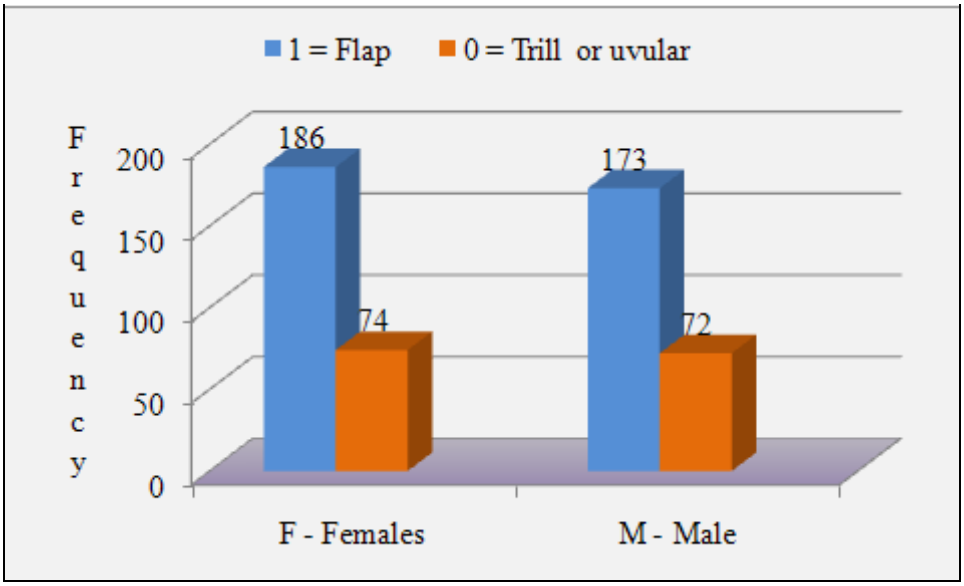


Figure 5: Frequency of initial flap [ɾ] = 1 by sex

In terms of the percentage of recorded tokens for females and males respectively as reflected in Figure 6, the performing likelihood of initial flap by both sexes shows a very slight difference of 1% favouring the females.

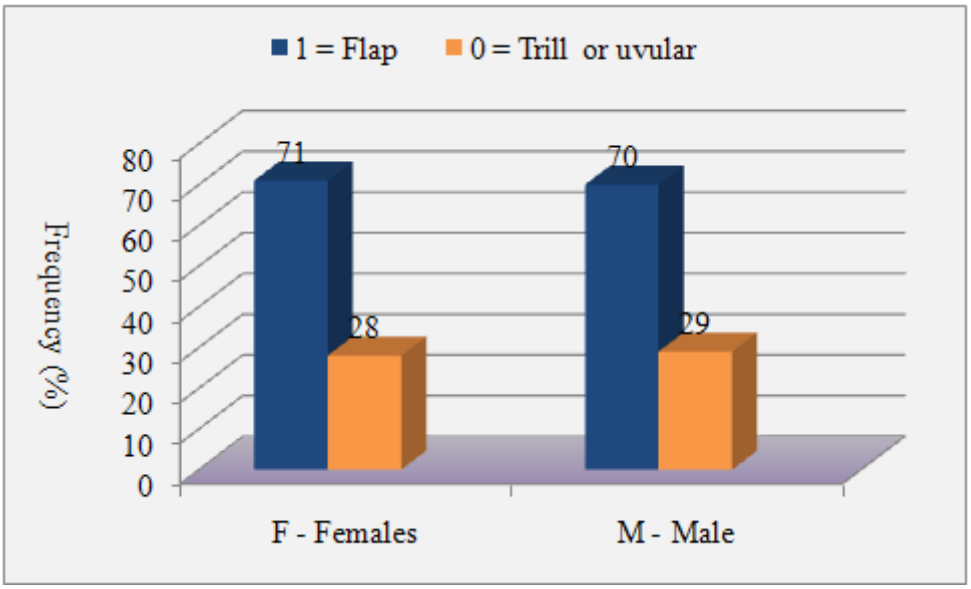


Figure 6: Percentage of initial flap realization [ɾ] = 1 by sex

The last co-occurrence examined is between the initial realization of flap and speakers of different generations. Figure 7 below shows the results.

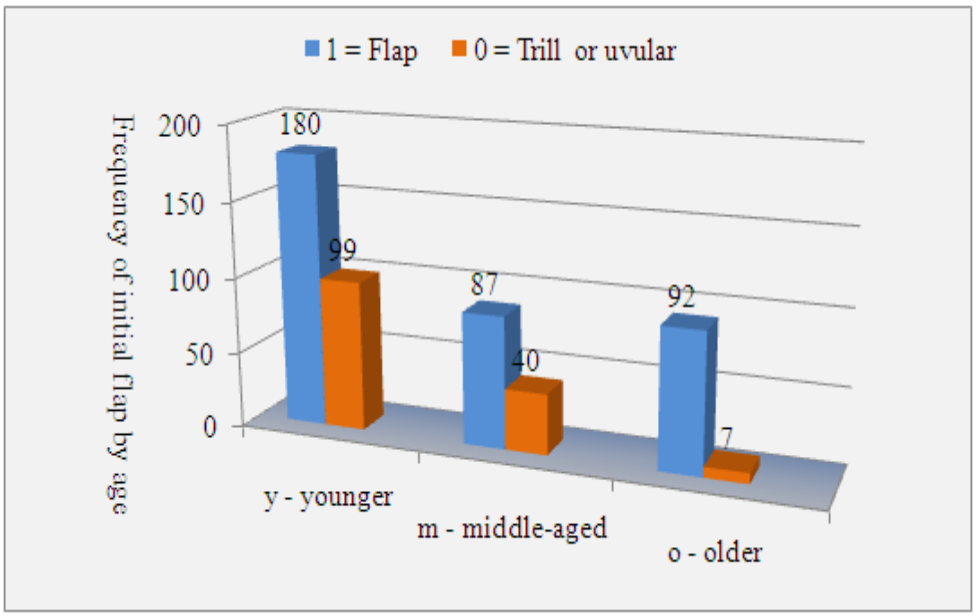


Figure 7: Frequency of initial flap [ɾ] = 1, stratification by age

According to frequency of tokens recorded in each group age, results of initial realization of flap show 180 for younger, 87 for middle age and 92 for older speakers. Figure 8 below shows the

percentages of frequencies of initial flap according to age group, as follows, from least to most: y (64%) > m (68%) > o (92%), respectively.

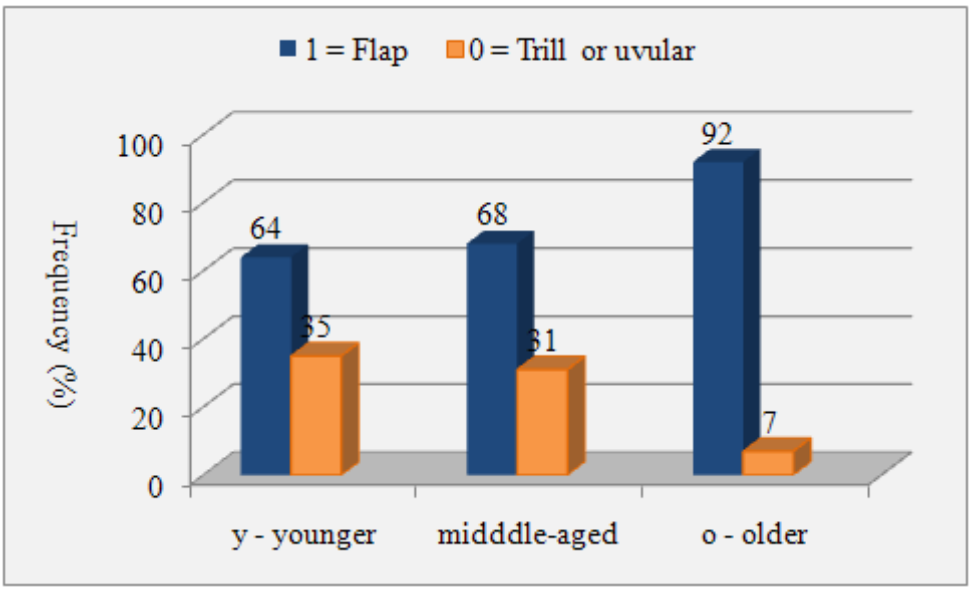


Figure 8: Percentages of initial flap realization [ɾ] = 1 by age

The *step-up/step-down* analysis in VARBRUL enables me to test if a given factor group is significantly relevant to be considered a pattern of variation (Bayley 2002:128). In the following analysis I will report on the significances of the differences regarding the co-occurrence of initial flap and the selected factor groups for this study, namely: residence, education, sex (and gender) and age.

In the following VARBRUL analysis I took the default application value to be '0', i.e., the results presented below assess the significance of the trill or uvular variant. Since there are only two values '0' and '1', this is the same as assessing the significance of the flapped variant.

The VARBRUL results demonstrate that the best stepping-up and the best stepping-down runs are #6 and #17, respectively. Table 33 shows the best stepping-up and best stepping-down runs for the analysis of Case One.

Best stepping up run: #6	Best stepping down run: #17
Input 0.742 Group #1 -- S: 0.457, U: 0.654 Group #4 -- y: 0.419, m: 0.418, o: 0.793 Log likelihood = -282.673 Significance = 0.012	Input 0.742 Group #1 -- S: 0.457, U: 0.654 Group #4 -- y: 0.419, m: 0.418, o: 0.793 Log likelihood = -282.673 Significance = 0.288

Table 33: Best stepping-up and best stepping-down runs of Case One

Table 33 shows comparable figures for initial flap realization. Thus, education (P, H) and gender (F, M) are not significant in the choice of one variant over the other. However, the other two social factors tested in this study, namely place of residence (S, U) and age (y, m, and o) are significant. The table also shows that the overall probability of occurrence of the initial flap is 74%, considering the input value rounded to two decimal places.

Considering residence, this implies that residing in urban areas of Maputo favors initial flap, at .65, while suburban speakers disfavor initial flap at .45.

In terms of age, the table shows closely parallel results for initial flap realization by younger and middle aged speakers, at .41 in both. Therefore, although this correlation between age and initial flap realization is apparent, it is not less plausible to assume that in this case being younger and middle-aged speakers disfavors the occurrence of initial flap among speakers in Maputo City. According to Bayley (2002:118) factor weights under .50 disfavor the dependent variable. However, quite the contrary, the result of being older is almost twice as high: that is, the older are much more likely than younger and middle aged speakers to use flap at the initial position, at .79. According to Chambers (2002:355) a variant representing a change in real time is marked by increasing frequency down the age scale, i.e., it will show from least to most this pattern of frequency: oldest > middle > youngest. From this I have to consider the initial flap realization in Maputo City as a stable variant and as an established class-marker (Chambers 2002:359).

According to the results of the best stepping-up run in Case One the co-occurring factors that were found to constrain the flap realization of phoneme /ʀ/ at the initial position are residence and age of the speakers, as indicated in Table 34 below.

Factor		Weight	Input	Range	Log likelihood	Significance
Residence	S – Suburban area	.45	.74	19	-228.673	0.012 p < 0.05
	U – Urban area	.65				
Age	y – younger	.41		37		
	m – middle-aged	.41				
	o – older	.79				

Table 34: Constraints on initial flap [ɾ]

Regarding the range of the two co-occurring external factors that were considered significant (residence and age), age has the strongest effect of 37 if compared to that of residence, at 19. Results of Table 34 show clear differences on initial flap realization between different speaker's generations. In fact, my apparent time data in this study, consisting of weights of initial flap realization [ɾ] by the three age groups (Table 35), show a strong increase from the middle age and younger speakers (.41) to the older speakers (.79).

The place of residence was another social factor with the strongest effect on the dependent variable. In fact, according to whether the speaker is from suburban or urban area, the weights indicate that flap realization is highly favoured by urban contexts at .65, and disfavoured in suburban contexts at .45. The run also shows that the two factors that were included in the analysis (Groups 1 and 4) have a statistically significant effect ($p < 0.05$ on the occurrence of the dependent variable). According to Robinson (2001:27) and Bayley (2002:129) figures that are close or above the $p < .05$ level are significant for the analysis. In other words, in the context of this analysis, this means that age and residence correlate positively with the occurrence of initial flap realization.

Figure 9 is a cross-tabulation (Robinson, J., and et al. 2001:21) of initial flap [ɾ] by age and residence. The results allow comparing the suburban with the urban sociolinguistic patterns of initial flap [ɾ], the same variable distribution across different age groups. In fact, the results in figure below denounce clearly a direct causal relationship between age and initial flap realization. In the two residential areas, namely urban and suburban areas, sociolinguistic

patterns of initial flap realization by speakers of different age groups show a remarkable age effect. The highest values are found among older speakers in both residential areas. Middle age and younger speakers generally show lower proportions. In urban areas the general distribution of flap accrding to age follows the pattern, from least to most: y (76%) > m (82%) > o (94%), while in suburban areas in the same order, i.e. from least to most, the patern is y (64%) > m (62%) > o (92%), respectively.

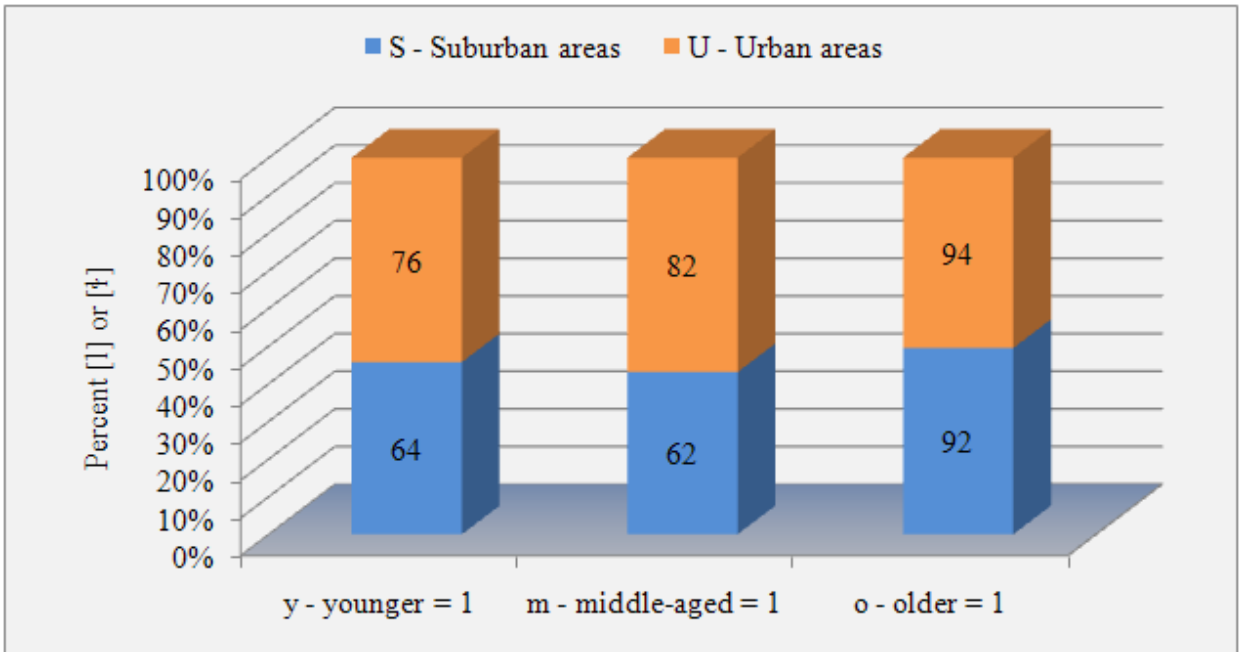


Figure 9: Cross tabulation percentage of initial flap [ɾ] by age and residence

5.2 Case Two: Middle flap realization

1 = Middle Flap [ɾ] and 0 = Middle uvular [ʁ] or trill [r]

The second variable to be considered in this analysis is middle realization of phoneme /ʀ/ represented by double "rr" in writing form and pronounced as flap [ɾ]. Figure 10 below is the first regression of middle "rr" as flap [ɾ] = 1, by residence in Case Two.

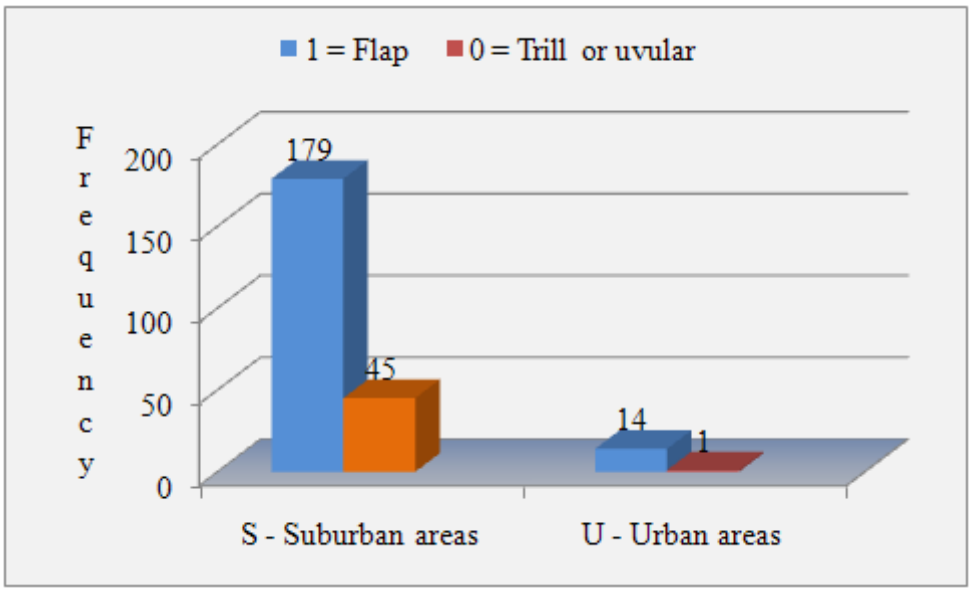


Figure 10: Realization of flap [ɾ] = 1 at the middle position by residence

Frequencies in Figure 9 show that middle realization of flap in suburban areas is at 179 and in urban ones is at14. In terms of comparative percentages of middle flap realizations, Figure 11 shows a clear powerful effect of urban areas on middle flap [ɾ] with 93%, while in suburban areas percentage reaches only 79%.

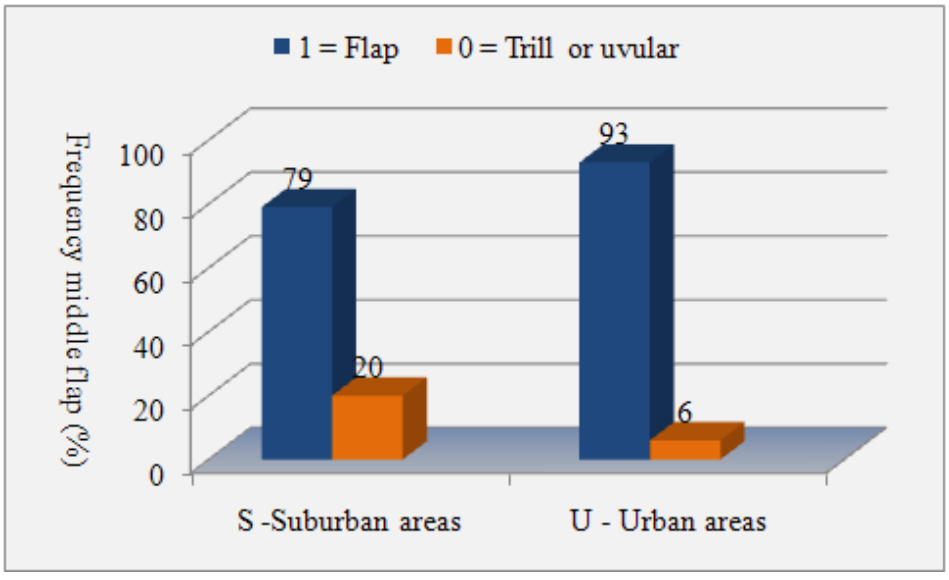


Figure 11: Percentage of middle flap realization by residence

Results of frequency in Figure 12 show the realization of flap [ɾ] at the middle position in a context of co-occurrence with education. The results from VARBRUL indicate that speakers with primary education use a flap veriant at similar levels of frequency to that of speakers with higher levels of education.

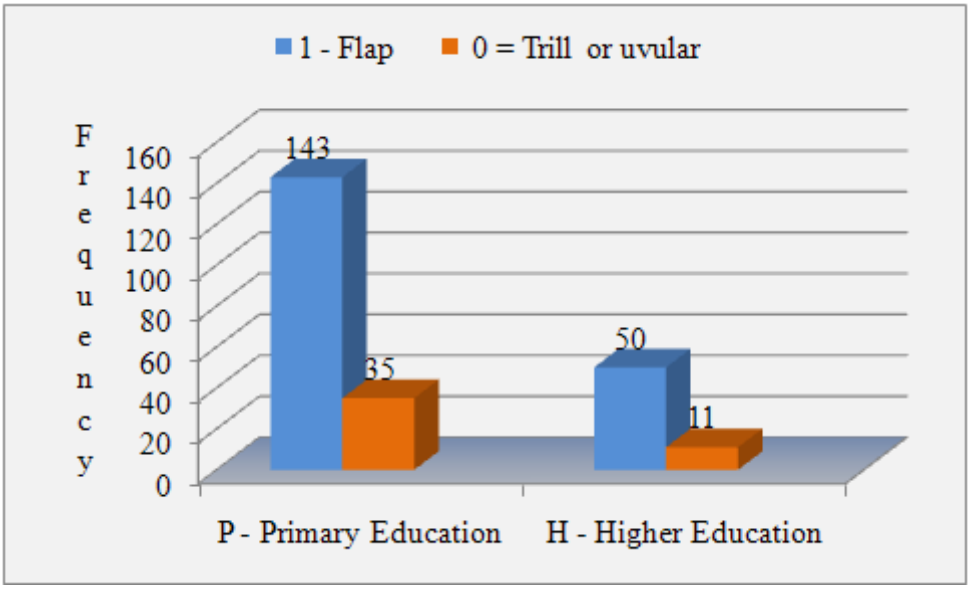


Figure 12: Frequency of middle flap by education

The following Figure 13 reflects clearly the same results in percentages. As can be seen clearly the difference of middle flap realization between speakers of different levels of education is of a minimum of 1%. These results are very close to that discussed previously in Case One.

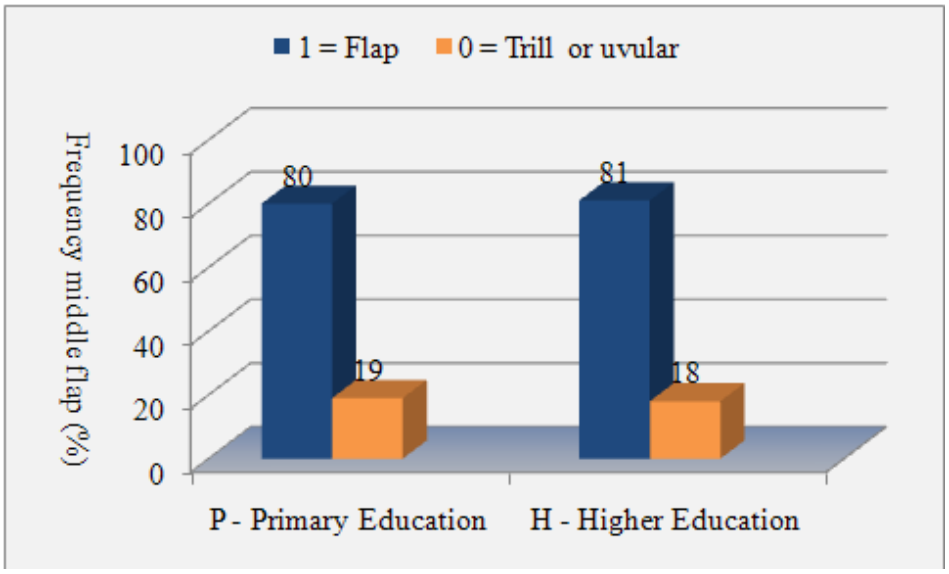


Figure 13: Percentages of middle flap by education

According to Labov (2001:263) sex differentiation as one of social factors is part of the explanation of apparent time linguistic change. Figure 14 below shows that are the men who indicate a relative tendency for high frequency of flap realization at 108, while the women produce middle flap at a relatively low levels of frequency of 85.

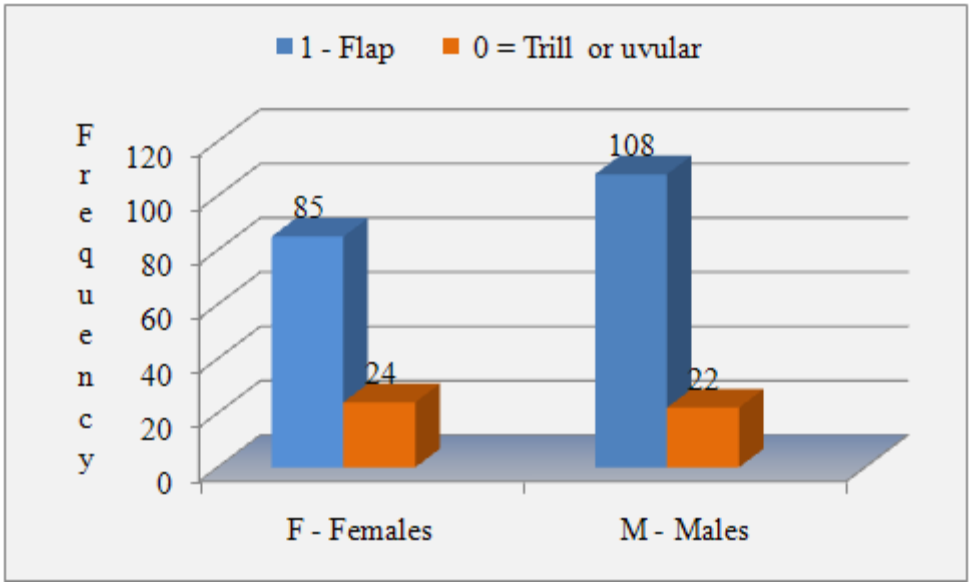


Figure 14: Frequency of medial flap realization by sex

The same results of middle flap realization by males and females speakers are here indicated in percentages. Figure 15 shows a weak effect of gender of the speaker in the realization of middle flap as demonstrated below.

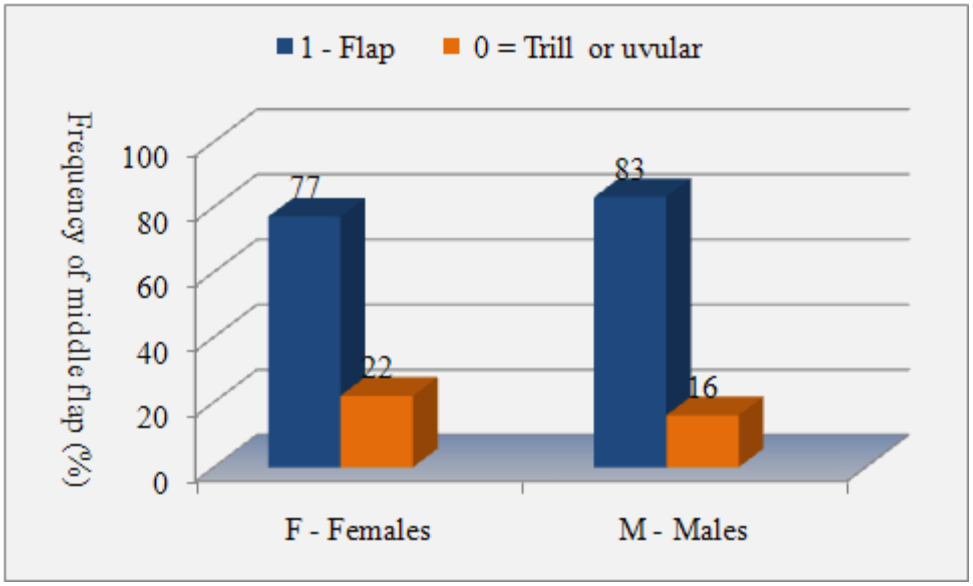


Figure 15: Percentage of medial flap by sex

Concerning apparent-time distribution of the medial flap based on age differentiation, Figure 16 and 17 of frequency and percentage respectively show a clear distinction between the younger and middle age speakers, on one side, and, on the other, the older groups. While the first group of younger and middle speakers indicates a very close parallelism of middle flap realization at 77% and 78%, older groups favouring middle flap realization at 92%. The general distribution of flap in regard to age follows the pattern, from least to most: y (77%) > m (78%) > o (92%).

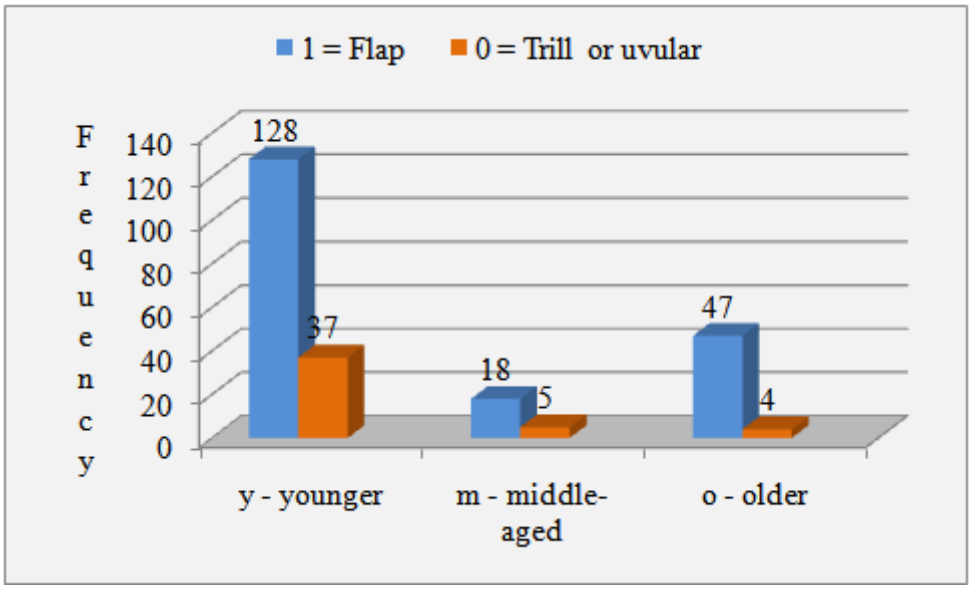


Figure 16: Frequency of middle flap realization by age

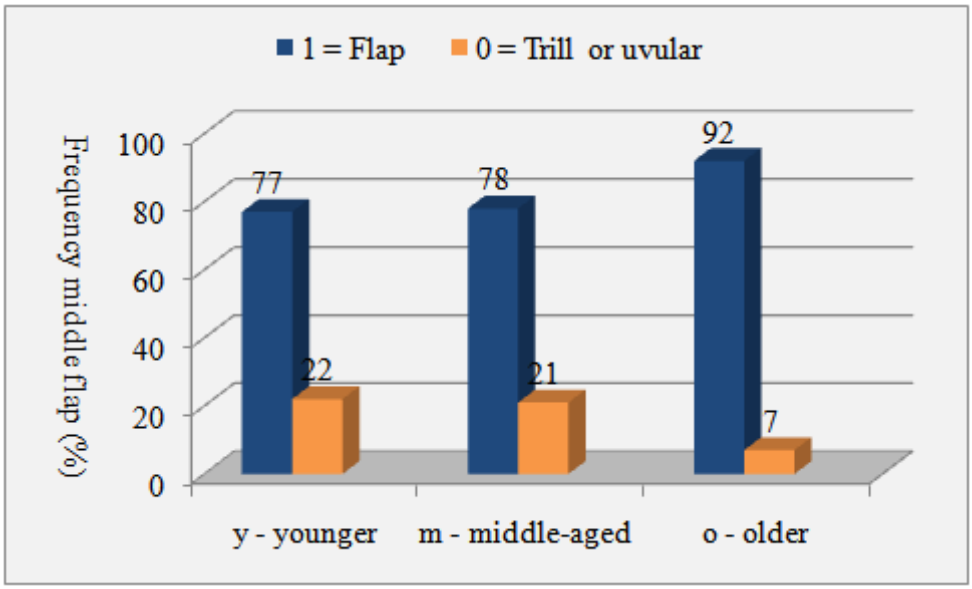


Figure 17: Percentage of medial flap by age

In order to determine the factor groups favouring flap realization in middle position I explore *step-up/step-down* analysis based on VARBRUL. In Case Two the results show that the best stepping-up and the best stepping-down runs are #5 and #17. In the following Table 35 are given the best stepping-up run and the best stepping down run for the analysis of Case Two.

Best stepping up run: # 5	Best stepping down run: #17
Input 0.818 Group # 4 -- y: 0.434, m: 0.444, o: 0.722 Log likelihood = -113.881 Significance = 0.044	Input 0.818 Group # 4 -- y: 0.434, m: 0.444, o: 0.722 Log likelihood = -113.881 Significance = 0.433

Table 35: Best stepping-up and best stepping down runs of Case Two

VARBRUL analysis shows that in all age groups the overall probability that /ʀ/ will occur as flap [ɾ] instead of uvular [ʁ] or trill [r] in middle position is at .81. The run shows that the single factor that was included in the analysis (i.e., the Group 4) has a statistically significant effect ($p < 0.05$ on the occurrence of the flap as the dependent variable. Table 35 above shows also a significant difference between the occurrences of the dependent variable according to whether younger, middle-aged, or older speakers appear in the factor group. Therefore, the factor weights show that flap occurrence is highly favoured among older speakers at .72, and disfavoured by middle-aged and younger speakers, at .44 and .43 respectively. The situation described above means that at the overall probability of 81% older speakers favour middle flap at 72%, while younger and middle speakers disfavour a middle flap at 43% and 44%.

The relative strength of the statistically effect, which is calculated by subtracting the lowest value from the highest of group 4, shows a range of 28. Table 36 below summarizes the constraints on middle flap [ɾ] in Case Two of my study.

Factor 4		Weight	Input	Range	Log likelihood	Significance
Age	y – younger	.43	.81	.28	-113.881	0.044 $p < 0.05$
	m – middle-aged	.44				
	o – older	.72				

Table 36: Constraints on flap in middle position

5.3 Case Three: Middle uvular or trill realization

1 = Middle uvular [ʁ] or trill [r] or 0 = Middle flap [ɾ]

My third linguistic variable tested in this analysis is middle flap [ɾ] realization as a uvular [ʁ] or trill [r]. Figure 18 is the first regression analysis of middle flap realization as a trill or uvular by residence in Case Three.

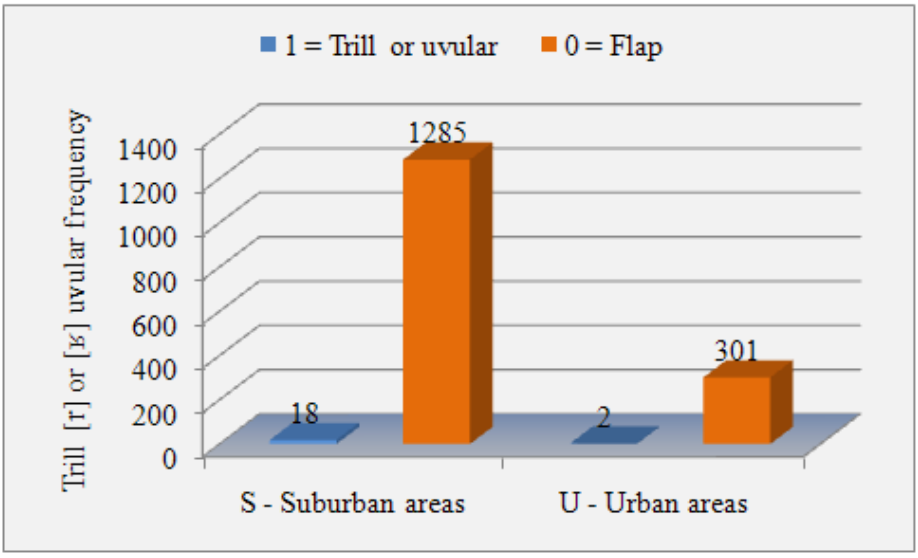


Figure 18: Frequency of middle trill or uvular by residence

The results of frequencies in VARBRUL analysis above combined with the percentages, as indicated in Figure 19 below, show in this case that urban speakers are more likely to produce a flap as trill or uvular at the middle position than speakers in suburban areas.

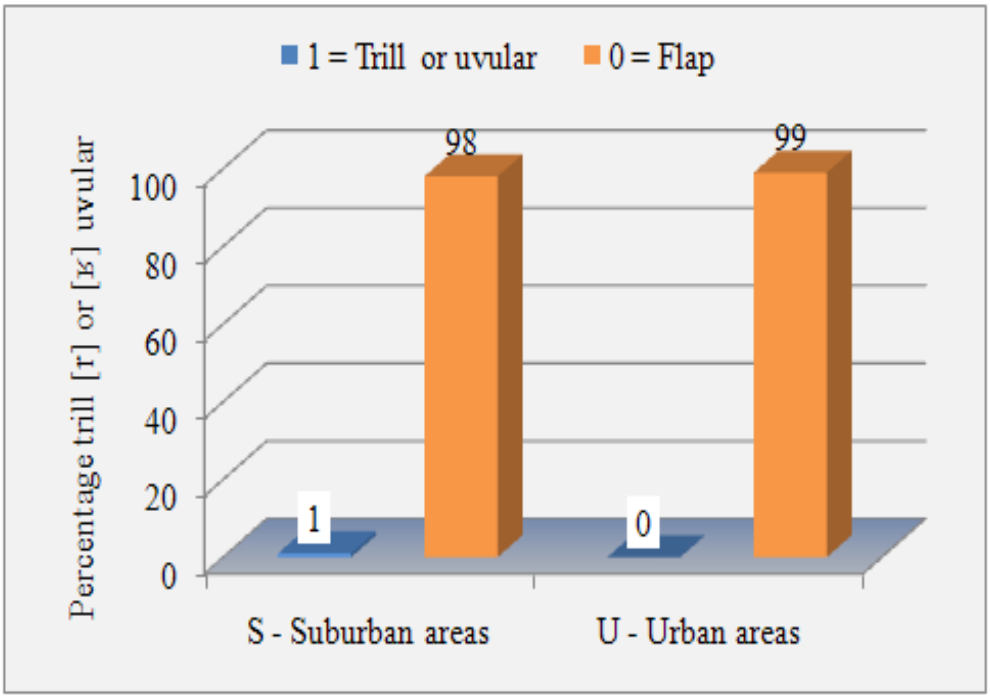


Figure 19: Percentages of middle trill or uvular by residence

In fact, according to the VARBRUL results, Figure 19 above shows that the overall chance that flap will occur as a trill or uvular at the middle position is of 99% in urban and of 98% in suburban areas, indicating a very slight difference of 1%.

In terms of speakers' level of education, results in Figure 20 show frequencies of 990 and 596 in trill or uvular realization at the middle position, by speakers with primary education and by speakers with higher level of education respectively. Results in percentage given in Figure 21 reflect parity in trill or uvular realization between lower and higher educated speakers at a higher degree of 98%.

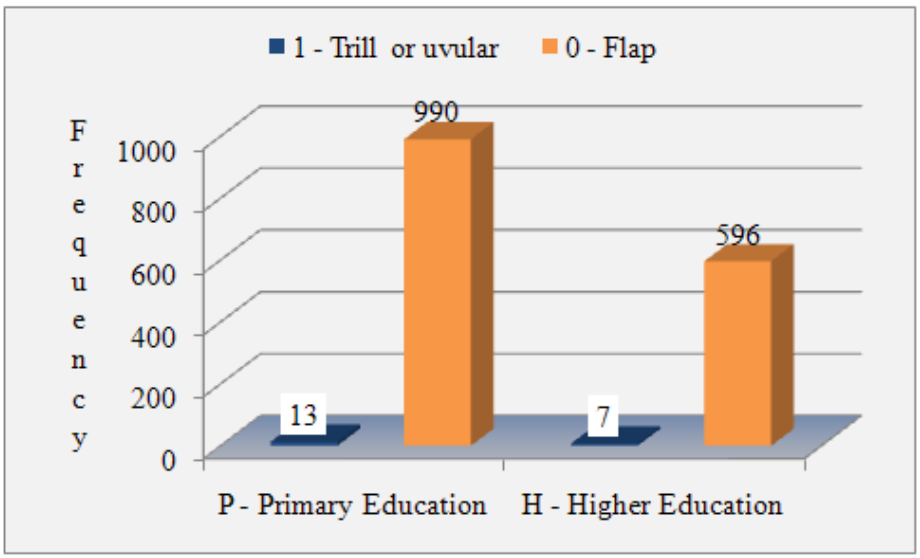


Figure 20: Frequency of middle trill or uvular by education

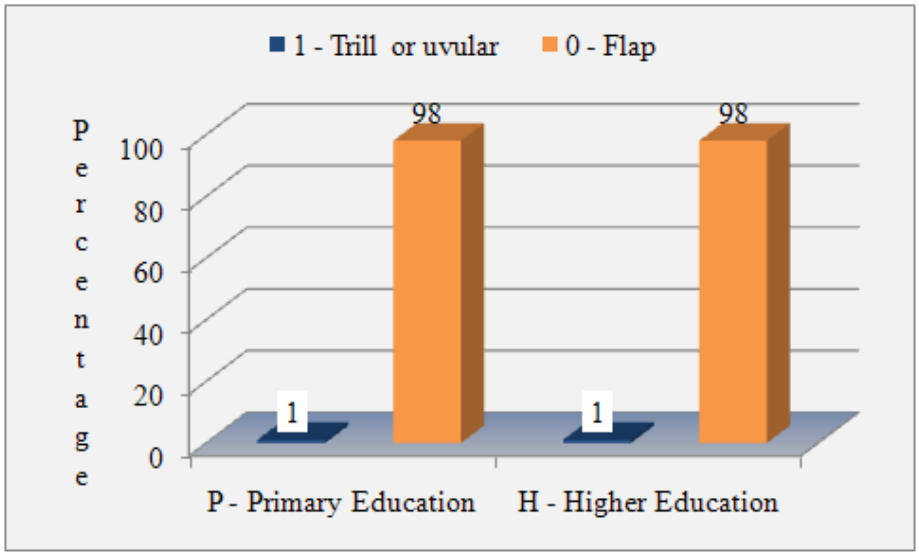


Figure 21: Percentages of middle trill or uvular by level of education

In terms of co-occurrence of trill or uvular with sex (and gender) VARBRUL frequencies in Figure 22 are also consistent indicating close results regardless to the sex of the speaker.

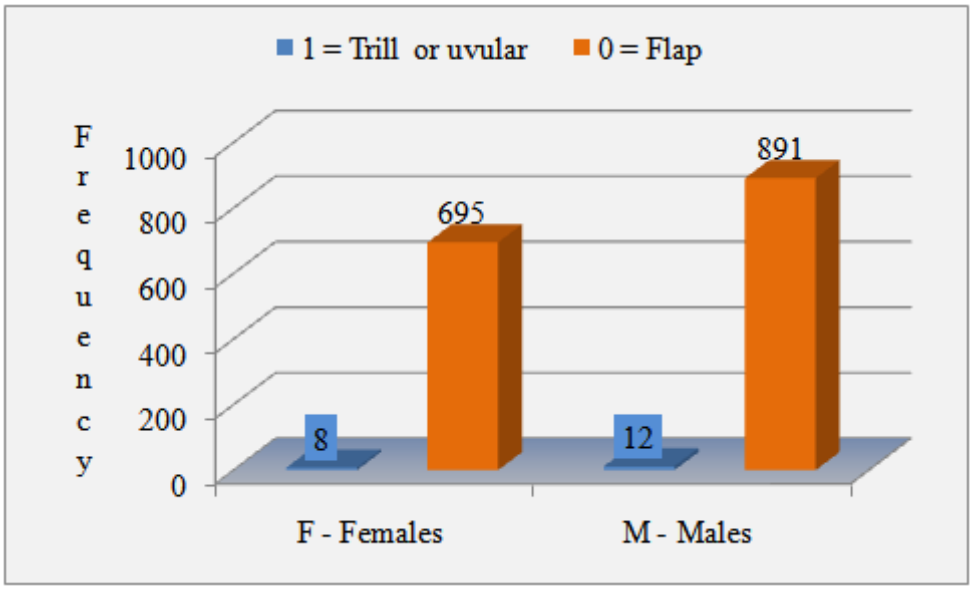


Figure 22: Frequency of trill or uvular by sex

The percentage results show more clearly that the realization of trill or uvular at the middle position instead of flap by males and females speakers is equal as demonstrated in Figure 23 below.

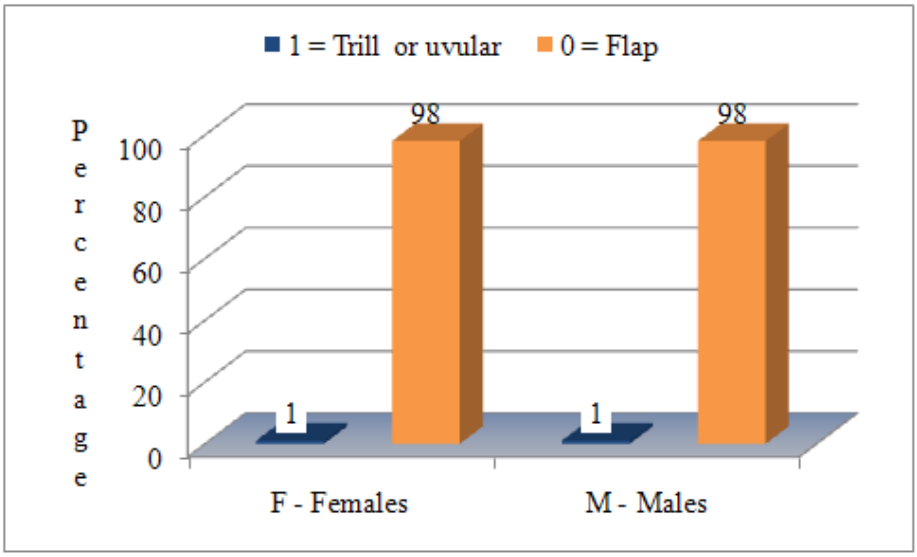


Figure 23: Percentages of middle trill or uvular by sex

The following figures 24 and 25 report results on middle trill or uvular co-occurrence with age in frequency and percentage. A combining interpretation of frequency and

percentage results allow me to conclude that age has a significant effect on the realization of flap as a trill or uvular at the middle position.

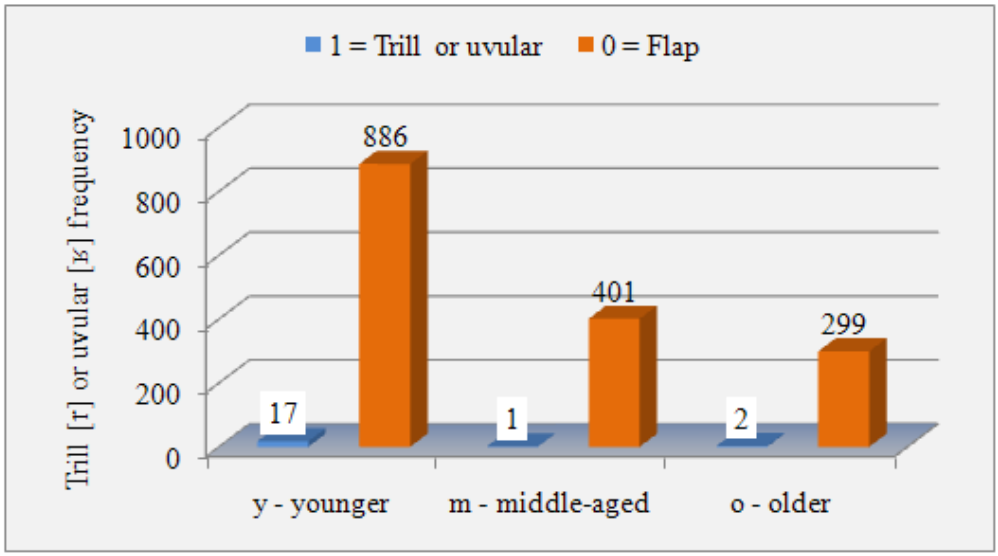


Figure 24: Frequency of trill or uvular by age

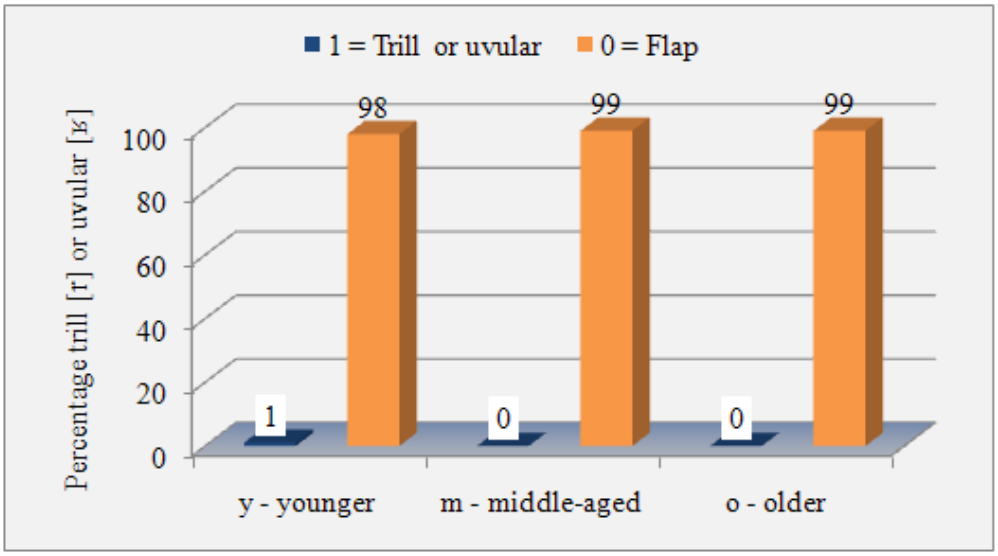


Figure 25: Percentage of trill or uvular by age

As shown in Figure 25 of percentages of trill or uvular realization in middle position, middle-age and older speakers show equal percentage of .99 while younger speakers indicate a realization of .98.

The *step-up/step-down* analysis based on VARBRUL in Case Three shows that the best stepping-up and the best stepping-down runs are #5 and #17. Table 17 gives the best stepping-up run and the best stepping down runs for the case in analysis.

Best stepping up run: # 5	Best stepping down run: #17
Input 0.009 Group # 4 -- y: 0.670, m: 0.209, o: 0.414 Log likelihood = -103.388 Significance = 0.016	Input 0.009 Group # 4 -- y: 0.670, m: 0.209, o: 0.414 Log likelihood = -103.388 Significance = 0.491

Table 37: Best stepping-up and best stepping-down runs of Case Three

In this run the overall probability that a medial trill or uvular realization will occur instead of a flap is at .00, and above the desired probability level (the $p < 0.05$ level) rounded two decimal places. This must be the clearest indication in this study that Mozambican Portuguese is experiencing variation and change in its phonological structure. The run shows that the single factor which was included in the analysis (in this case is the Group 4) has a statistically significant effect ($p< 0.05$ on the occurrence of the middle trill or uvular as the researching variable. It is also shown above the significant difference between the occurrences of the dependent variable according to age with younger groups favouring trill, at .67, whereas older, at .41, and middle-aged, at .20, disfavor the trill respectively. However, even considering the fact that youngest groups favor middle flap realization, the probability of this occurrence it is inexistent. Table 38 below summarizes the constraints on middle trill or uvular in Case Three.

Factor 4		Weight	Input	Range	Log likelihood	Significance
Age	y – youger	.67	00	46	-103.388	0.016 p < 0.05
	m – middle-aged	.20				
	o – older	.41				

Table 38: Constraints on trill in middle position

5.4 Case Four: Middle lateral or fricative

1 = Middle lateral [l] or fricative [ɬ] and 0 = Middle palatal [ʎ]

The last token file of this study focus on lateralization or fricatization of lateral palatal/ʎ/ at the middle position. The results of first run of GoldVarb in Figure 26 show a frequency of lateralization or fricatization of 397 in suburban areas and of 32 in urban areas.

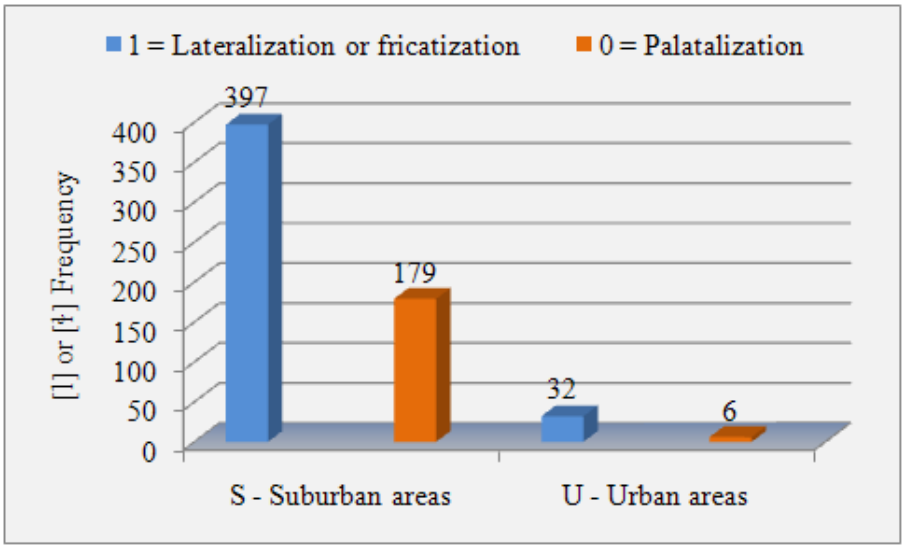


Figure 26: Frequency of middle lateralization or fricatization by residence

In terms of percentage Figure 27 also demonstrate a strong effect of area of residence, and the results indicate that urban speakers are likely to lateralize or fricatize at 84% and speakers from urban areas at 68%.

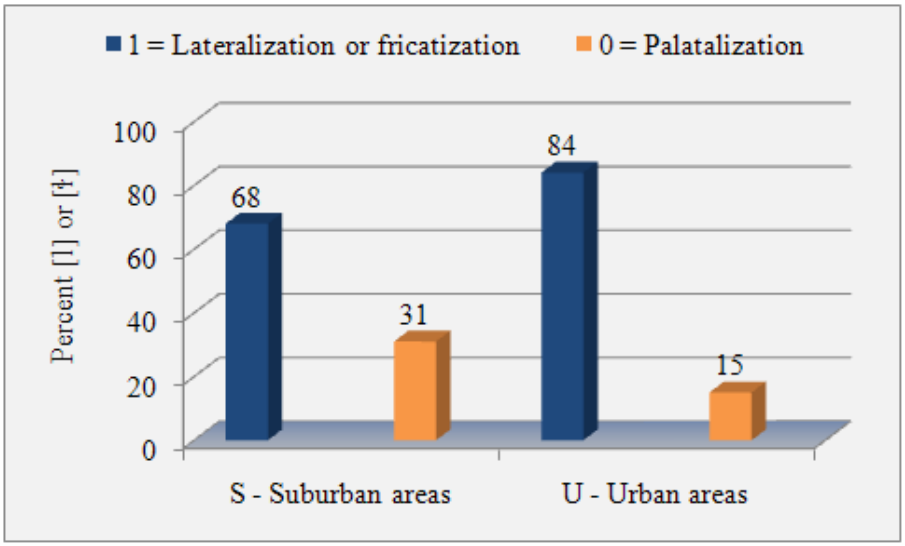


Figure 27: Percentage of lateralization or fricatization by residence

Regarding to the education a comparative analysis of results of requency in Figure 28 and percentages in Figure 29 reflects a greater tendency towards lateralization or fricatization according to residence. In fact, higher educated speakers lateralize or fricatize at a level of (72%) in relation to lower values of speakers with primary level of education (68%).

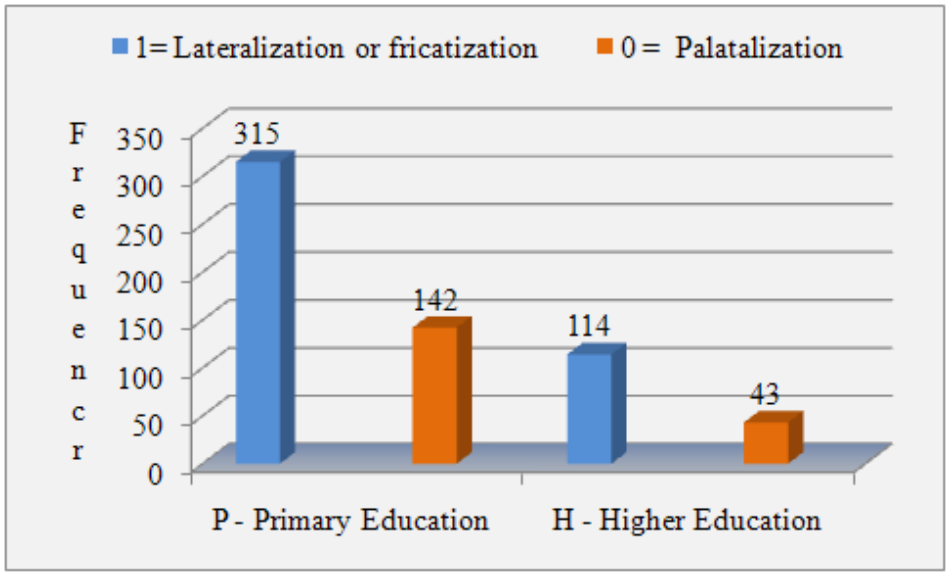


Figure 28: Frequency of lateralization/fricatization by education

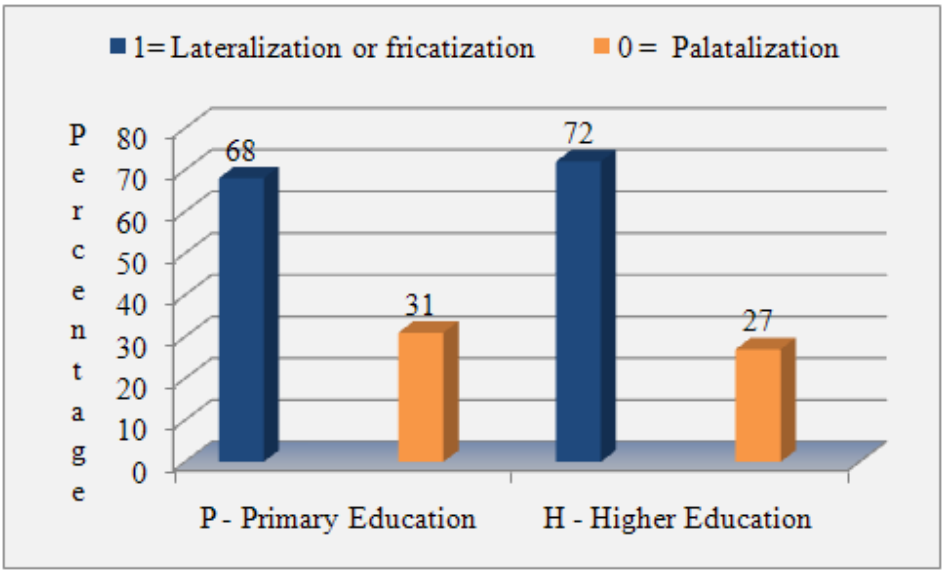


Figure 29: Percentage of lateralization or fricatization by education

The spreading of lateralization or fricatization with respect to sex (and gender) of the speaker is shown in Figure 30 and in Figure 31 below.

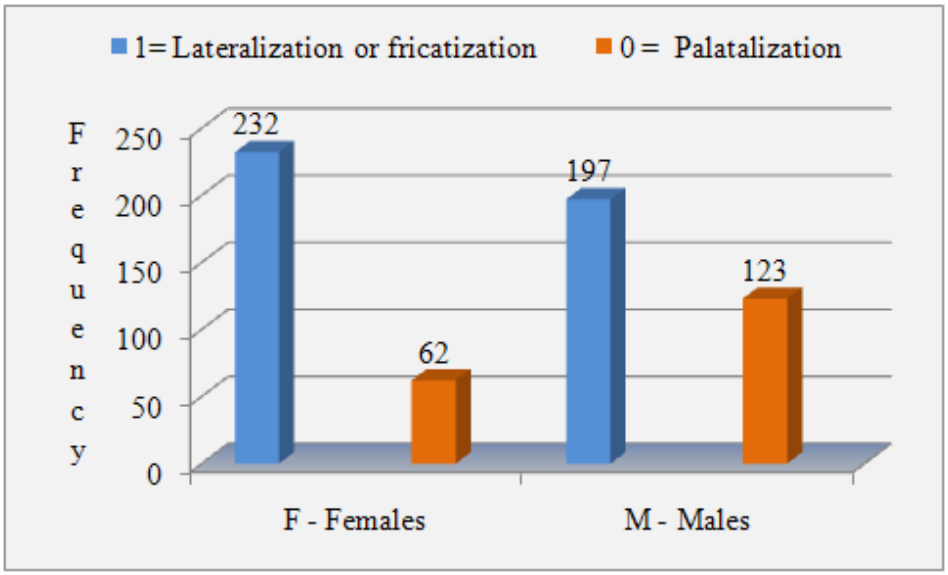


Figure 30: Effect of sex on lateralization or fricatization

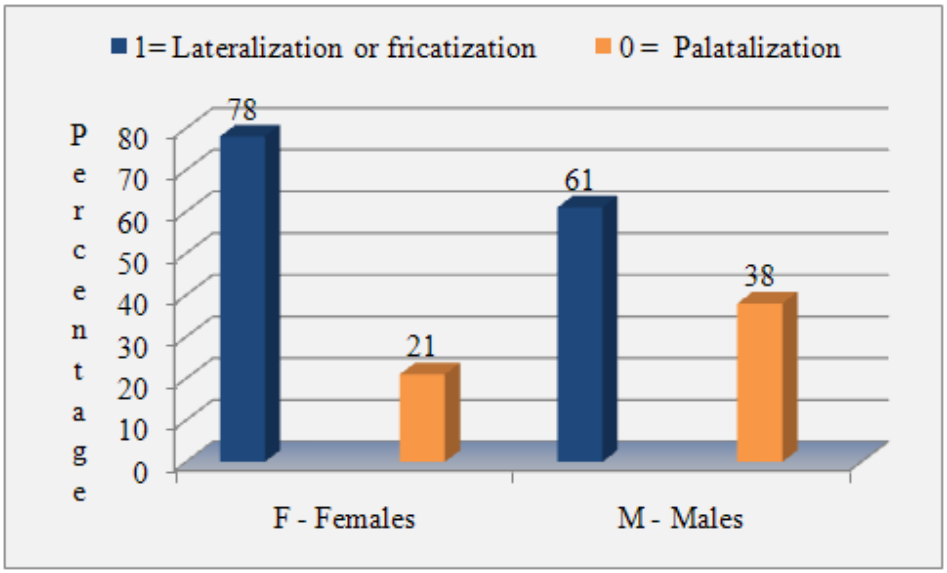


Figure 31: Percentage of the effect of sex on lateralization or fricatization

The combined analysis of frequency and percentages indicate strong sex (and gender) effect on lateralization or fricatization. Females show higher values of frequency and percentage of lateralization or fricatization (232 and 78%) than males (with 197 and 61%).

Age is another external factor tested in co-occurrence with middle lateralization or fricatization. The frequencies and percentages are represented in figures 32 and 33 respectively.

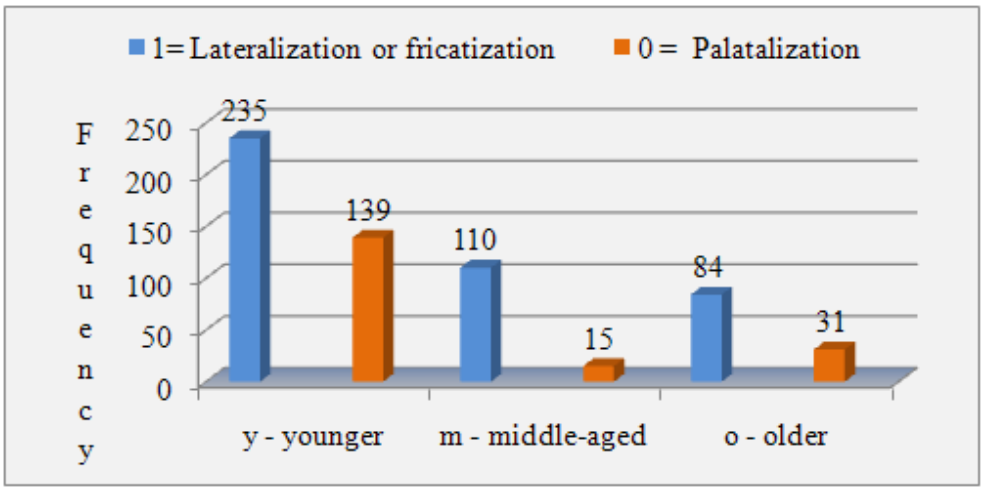


Figure 32: Frequency of age's effect on middle lateralization/fricatization

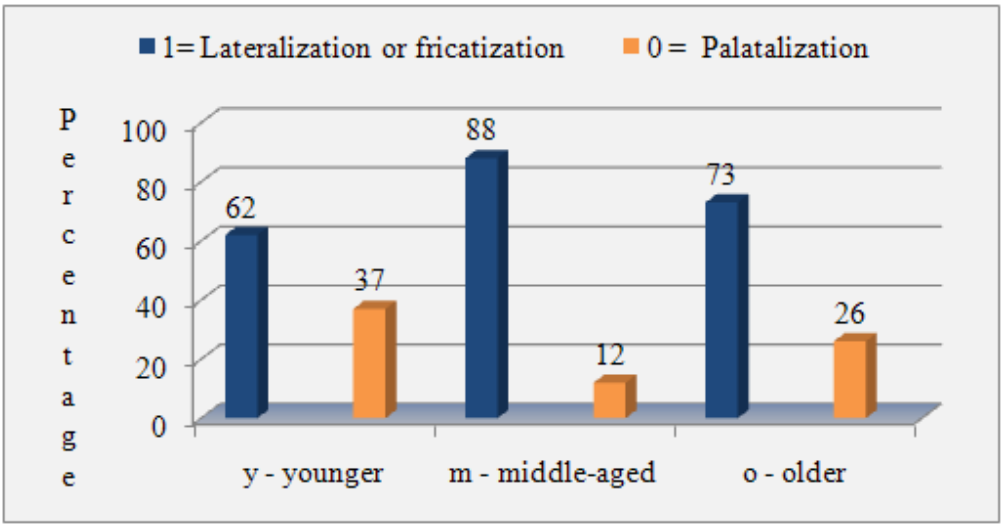


Figure 33: Percentage of age effect on middle lateralization or fricatization

Results of frequency and percentage of the interaction between lateralization or fricatization and age in figures above indicate the highest values for middle age groups, followed by the older and at the end are the younger speakers with lower values.

The following results based on *step-up/step-down* analysis in VARBRUL reports on the significances of the differences found previously regarding the co-occurrence of lateralization or fricatization and the following external factors: residence, education, sex (gender) and age. The VARBRUL results demonstrate that the best stepping-up and the best stepping-down runs are #10 and #13, respectively as can be seen in Table 39.

Best stepping up run: # 10	Best stepping down run: #13
Input 0.717	Input 0.717
Group # 2 -- P: 0.467, H: 0.594	Group # 2 -- P: 0.467, H: 0.594
Group # 3 -- F: 0.595, M: 0.413	Group # 3 -- F: 0.595, M: 0.413
Group # 4 -- y: 0.432, m: 0.703, o: 0.488	Group # 4 -- y: 0.432, m: 0.703, o: 0.488
Log likelihood = -353.009 Significance = 0.021	Log likelihood = -353.009 Significance = 0.243

Table 39: Best stepping-up run of Case Four

The results of the best stepping-up run in this case show that the co-occurring factors that were found to constrain the variation of lateralization or fricatization in middle position in Case Four are education, sex and age as summarized in Table 40 below.

Factor groups: 2, 3 and 4		Weight	Input	Range	Log likelihood	Significance
Education	Primary	.46	.71	12	-353.009	0.021 p < 0.05
	Higher	.59				
Sex	Females	.59		18		
	Males	.41				
Age	younger	.43		27		
	m-aged	.70				
	older	.48				

Table 40: Constraints on lateralization or fricatization in middle position

In regard to the range of the three external factors that were considered significant, age has the strongest effect. Findings of Table 40 show that younger and older groups disfavour lateralization or fricatization at (.43) and (48), while middle-aged speakers at (.70) favour lateralization or fricatization.

Among the analysed factors, education was selected as a significative social factor with the strongest effect on lateralization or fricatization. Speakers with higher levels of education favor lateralization or fricatization at (.59) while speakers with primary level disfavour middle lateralization or fricatization at (.46).

Sex (and gender) differentiation was also considered a significant factor, with females favouring lateralization or fricatization with an index of (.59). On the contrary, male speakers' disfavour lateralization or fricatization of palatal at the middle position with an index of (.41).

Figure 34represents the cross-tabulation function (Grid) between age and education. The results obtained allow to compare the percentage of age distribution of [l] or [ɬ], the sociolinguistic variables of this case, across speakers of different education background.

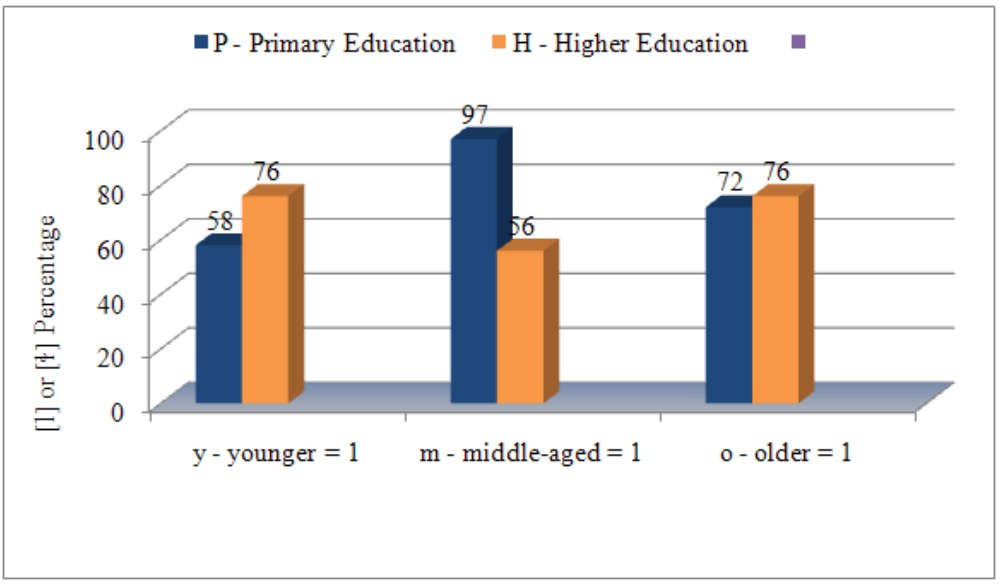


Figure 34: Cross-tabulation percentage of [l] or [ɬ] by education and age

Although the findings of my study can be partially testable by Chambers' Eight Rules of Dialect Acquisition, as detailed in chapter 4 of this study, it is worthwhile to make few remarks on the differences between the two studies. First, Chambers' eight principles are generalizations drawn from his observations of linguistic behaviour of six monolingual native speakers of the Canadian English (CE) dialects in their process of acquiring linguistic features of another English variety, namely the Southern England English (SEE) dialects. Second, the subjects of Chambers' study are not a random sample. Third, despite its workability in terms of "ratio between sample size and analytical returns", Chambers notes that the subjects of his study were few to be representative of the dialect acquirers, a strong recommendation for future application of his principles. Fourth, Chambers' youngsters were interviewed two years after they moved to southern England. Fifth, Chambers' methodology and data included multiple techniques and outputs among discussions, tape listening, identification of object and reading. In my study, on the other hand, the 20 subjects are part of a random sample of almost 100 bilingual speakers (92% are LB/L1), being therefore, and considered representative of the MP speech community of city of Maputo. From this view point, my study comprises only oral data and it establishes Standard 4 as a required minimal proficiency of the participants in their L2. I now analyse how my data can be related to Chambers' postulations and theoretical positions outlined in chapter 3 regarding dialect acquisition.

My study confirms a correlation between the variation patterns that characterise the variant allophonic distribution and other ongoing process of variation of Mozambique Portuguese and related social patterns. First, considering the assumption that different linguistic forms among generations can be associated with diachronic change, I propose that the linguistic behaviour of each generation of Mozambican speakers of Portuguese reflects a specific language stage of the Mozambican Portuguese of city of Maputo. Younger groups are using new forms of Portuguese that gradually will replace the linguistic patterns of older groups. Although Chambers in Rule 4 is cautious about associating early acquirers by age group, the results of Case Three of this study are the most relevant in this regard.

In fact the figures in the best stepping-up run of this case shows that the overall probability for alveolarization or uvularization of flapped phoneme /ʀ/in middle position is .001 (rounded to two decimal places) and a statistically significant effect ($p < 0.05$) on the occurrence of the dependent variable. Age is the strongest social factor in alveolarization or uvularization in Mozambican Portuguese, indicating clearly differences between older and middle-aged speakers and younger generations. Younger speakers favour and middle-aged and older speaker's disfavour the alveolarization or uvularization in the middle position.

The findings in all cases analysed in this study confirm partially that all factors tested in this study, such as education, gender and place of residence influence the tendency of the speakers of Mozambican Portuguese to the realization of flapped /ʀ/ instead of a trilled in initial and middle positions. They also indicate that lateralization or fricatization is influenced more by age if compared to gender and education of the speakers. Despite the strong claim of the authors of the *PPOM* corpus that good proficiency in Portuguese is related to education level of the speakers, in terms of phonological acquisition my results surprisingly show that education has the weakest effect. In Case Four, for example, results indicate that speakers with higher levels of education favor lateralization or fricatization at (.59) while speakers with primary level disfavour middle lateralization or fricatization at (.46).

In terms of residence, the results of this study do not corroborate the *PPOM* authors implicit claim that residents in urban areas will perform much better than their counterparts of the suburban areas. In fact, findings in Case One show that, in terms of acquisition of

phonological features, residing in urban areas of Maputo favors initial flap, at .65, while suburban speakers disfavor initial flap at .45.

Table 42 below illustrate in summary that the sociolinguistic patterns demonstrated by flap [ɾ], trill [r] or uvular [ʁ] and lateral [l] or fricative [ɬ] are strongly affected by age. Residence and sex play a moderate role, while education represents the weakest effect.

Factor groups		Case One	Case Two	Case Three	Case Four
		Flap = 1	Flap = 1	Velar or uvular = 1	Lateralization or fricatization = 1
1	Residence	19	-	-	-
2	Education	-	-	-	12
3	Sex	-	-	-	18
4	Age	37	28	46	27

Table 441: Summary of the strength of each statistically significant effect of social variables in percentages.

After all runs, a comparative analysis using a multivariate function of GoldVarb 2001 confirms Chambers' (1998) principle 4 that acquisition of complex rules and new phonological features splits into age groups. The findings are also consistent with principle 6 based on the theory of language diffusion. More important is that the comparison of different apparent-time data clearly indicated that alveolarization or uvularization of phoneme /ʀ/ in medial position is a current change in progress in Mozambican Portuguese.

CHAPTER 6
CONCLUSION

The main goal of this study was to determine the extent to which Chambers' Eight Principles of Dialect Acquisition can be related with my data analyzing the acquisition of some phonological features by speakers of Portuguese in city of Maputo. The results of this study show that there is a clear correlation between the variant allophonic distribution of the socio-phonetic variables /ʀ/ and /ʎ/ and related social factors. The results indicate age of the speakers as the social factor with strongest effect. The statistical analysis did not show evident effect of all other social factors tested namely residence, sex and education. Therefore, another crucial dimension to explore in future research will be the effects of language contact, considering the fact that the majority of Mozambique Portuguese speakers are bilingual. Such a study will help to determine how language contact in Mozambique is modifying the Portuguese.

I also believe that some of the findings of this study will also contribute to a better understanding of the processes of language acquisition and variation, and in this way, it will assist in producing language materials and methods. From the sociolinguistic point of view, this study can make contributions to understanding social stratification in southern Africa.

REFERENCES

Appel, R. & Muysken, P. 1987. *Language Contact and Bilingualism*. London: Arnold.

Bokamba, E. G. 1983. The Africanization of English. In *The Other Tongue – English Across Cultures.* B.B. Kachru, Ed. Oxford: Pergamo Press. 77-98.

Cardeira, E. 2006. *História do Português*. Lisboa: Editorial Caminho.

Castro, I. 2006. *Introdução à História do Português*. Lisboa: Edições Colibri.

Chambers. J. K. 1995. *Sociolinguistic Theory: Linguistic Variation and its Social Significance*. Oxford: Blackwell.

Chambers, J. K. 1998. Dialect Acquisition. In *The Sociolinguistics Reader: Multilingualism and Variation*. P. Trudgill & J. Cheshire, Eds. London: Arnold. 145-178.

Chambers, J. K. & Trudgill, P. 1980. *Dialectology*. Cambridge: Cambridge University Press.

Chambers, J. K., Trudgill, P. & Schilling-Estes, N. 2002. *The Handbook of Language Variation and Change*. Malden, MA: Blackwell Publishers Inc.

Cruz-Ferreira, Madalena 1999. Portuguese (European). In *Handbook of the International Phonetic Association*. Cambridge: Cambridge Unicersity Press.

de Araújo, M. G. M. n.d. *Espaço Urbano Demograficamente Multifacetado: As cidades de Maputo e da Matola.* [Research paper]. [Online]. Available: www.apdemografia.pt/pdf/ManuelAraujo.pdf [2010, April 24].

Feagin, C. 2002. Entering the Community: Fieldwork. In *The Handbook of Language Variation and Change*. J. K.Chambers, P. Trudgill & N.S. Estes, Eds. Malden, MA: Blackwell Publishers Inc. 20-39.

Ferguson, C. 1971. *Language Structure and Language Use*. Stanford, CA: Stanford University Press.

Ferguson, C. 1983. Foreword. In *The Other Tongue – English Across Cultures.* B.B. Kachru, Ed. Oxford: Pergamo Press. vii-xi.

Firmino, G. 2002. *A "Questão Linguística" na África Pós-Colonial: o Caso do Português e das Línguas Autóctones em Moçambique*. Maputo: PROMÉDIA.

Fought, C. 1999. A Majority Sound Change in a Minority Community: /u/-fronting in Chicano English. *Journal of Sociolinguistics*. 3(1): 5-23.

Freeman, D. L-. & Long, M. H. 1991. *An Introduction to Second Language Acquisition Research*. London: Longman.

Gonçalves, P. 1996. *Português de Moçambique: uma Variedade em Formação*.Maputo: IUEM.

Gonçalves, P. 2003. A Nativização da Língua Portuguesa em Sociedades Africanas Pós-coloniais: o Caso de Moçambique. *A Língua Portuguesa, Actas dos IX Cursos Internacionais de Verão da Cascais*. Vol. 2. Cascais, Portugal: Câmara Municipal de Cascais. 47-58.

Guitart, J. M. 2005. Sociophonetic Knowledge of Spanish and Control of Style. In *Selected Proceedings of the Second Workshop on Spanish Sociolinguistics*. [Paper #1137]. L. Sayahi & M. Westmoreland, Eds. Somerville, MA: Cascadilla Proceedings Project. 16-23. [Online]. Available: http://www.lingref.com [2010, April 24].

Gumperz, J. 1972. The Speech Community. In *Language and Social Context:* S*elected Readings.* P.P. Giglioli, Ed. Harmondsworth: Penguin Books. 219-231.

Hatton, J., Couto, M., & Oglethorpe, J. 2001. *Biodiversity and War: A Case Study of Mozambique*. Washington, D.C.: Biodiversity Support Program.

Hazen, K. 1998. The Birth of a Variant: Evidence for a Tripartite Negative Past *be* Paradigm. *Language Variation and Change.* 10: 221-224.

Instituto Nacional de Estatística. 2007. Resultados Preliminares: 3ºCenso Geral da População e Habitação. [Online]. Available: http://www.ine.gov.mz [2009, July 2].

Kachru, B. B. Ed. 1983. *The Other Tongue – English Across Cultures.* Oxford: Pergamo Press.

Kirk, J. 1994. Using VARBRUL for Studying Modal Auxiliary Verbs? *Hermes, Journal of Linguistics n*o.13 - 1994:71-86. [Oline]. Available: http://download1.hermes.asb.dk/archive/FreeH/H13_07.pdf [2009, August 20].

Labov, W. 1966. *The Social Stratification of English in New York City*. Washington, D.C.: Centre for Applied Linguistics.

Labov, W. Ed. 1980. *Locating Language in Time and Space*. New York: Academic Press.

Labov, W. 2001. *Principles of Linguistic Change.* Vol. 2. Malden, MA: Blackwell Publishers Inc.

Ladefoged, P. 1993. *A Course in Phonetics.* 3rd ed. Fort Worth, TX: Harcourt College.

Lass, R. 1984. *Phonology: An Introduction to Basic Concepts.* Cambridge: Cambridge University Press.

Lass, R. 2004. South African English. In *Legacies of Colonial English*. R. Hickey, Ed. Cambridge: Cambridge University Press. 363-386.

McEntegart, D. & Le Page, R.B. 1982. An Appraisal of Statistical Techniques used in the Sociolinguistic Survey of Multilingual Communities. In *Sociolinguistic Variation in Speech Communities*. S. Romaine, Ed. London: Edward Arnold. 105-124.

Martins, M. R. D. 1992. *Ouvir Falar: Introdução à Fonética do Português*. 2nd ed. Lisboa: Editorial Caminho.

Mateus, M. H. M., Brito, A. M., Duarte, I. S. & Faria, I. H. 1983. *Gramática da Língua Portuguesa.* Coimbra: Livraria Almedina.

Mendonça, F. 1988. *Literatura Moçambicana: a História e as Escritas*. Maputo: Faculdade de Letras e Núcleo Editorial.

Mesthrie, R. 1992. *English in Language Shift: The History, Structure and Sociolinguistics of South African Indian English*. Johannesburg: Witwatersrand University Press.

Morgan, T. A. 2006. On the Teaching of Spanish Pronunciation. In *Selected Proceedings of the 2nd Conference on Laboratory Approaches to Spanish Phonetis and Phonology*. [Paper #1330]. M. Díaz-Campos, Ed. Somerville, MA: Cascadilla Proceedings Project. 119-130. [Online]. Available: http://www.lingref.com [2010, April 24].

Moulton, W. G. 1962. *The Sounds of English and German*. Chicago: The University of Chicago Press.

NELIMO. 1989. *Relatório do I Seminário sobre a Padronização da Ortografia de Línguas Moçambicanas.* Maputo: INDE-UEM/NELIMO.

Ngunga, A. 2004. *Introdução à Linguística Bantu*. Maputo: IUEM.

Platt, J., Weber, H., & Ho, M. L. 984. *The New Englishes.* London: Routledge and Kegan Paul.

Roach, P. 2000. *English Phonetics and Phonology*. Cambridge: Cambridge UniversityPress.

Robinson, J., Lawrence, H. & Tgliamonte, S. 2001. *Goldvarb 2001: A Multivariate Analysis Application for Windows.*[Online]. Available: http://www.york.ac.uk/depts/lang/webstuff/goldvarb/manual/manualOct2001.html [2004, September 18].

Romaine, S. Ed. 1982. *SociolinguisticVariation in Speech Communities*. London: Edward Arnold.

Sitoe, B. & Ngunga, A. Eds. 2000. *Relatório do II Seminário sobre a Padronização da Ortografia de Línguas Moçambicanas*. Maputo: Empresa Moderna.

Stroud, C. & Gonçalves, P. Eds. 1997. *Panorama do Português Oral do Maputo*. Maputo: INDE.

Sureck-Clark, C. 1998. Dialect Acquisition and Prestige in Brazilian Portuguese. MA Dissertation. University of Pennsylvania, Pennsylvania. (Unpublished).

Tarallo, F. 1997. *A Pesquisa Sociolinguística*. São Paulo: Editores Ática & Editora Afiliada.

Trudgill, P. & Cheshire, J. Eds. 1998. *The Sociolinguistics Reader: Multilingualism and Variation*. London: Arnold.

Walters, K. 1988. Dialectology. In *Linguistics: The Cambridge Survey*. F. Newmeyer, Ed. Cambridge: Cambridge University Press. 119-139.

Wardhaugh, R. 1992. *An Introduction to Sociolinguistics.* 2nd ed. Oxford: Blackwell.

APPENDICES

APPENDIX 1: CASE ONE – DISTRIBUTION OF TOKENS

CASE ONE

1 = Initial flap [ɾ]

or

0 = Initial trill [r] or uvular [ʁ]

Identification of the participant: MR12RAF

Type of data: Recorded Interview (digitalized)

Total of tokens: 18

String code: (1SPFy

Initial /ʁ/ = Trill or uvular

			T/Tokens	Flap	N/Flap
Speech No.	**No.**	**Line**	**18**	**1**	**0**
18. RAF:	1	os rapazes possam começar com dezassete dezoito		/	
30. RAF:	3	que podem servir à mesa já começam a preparar a roupa - não		/	
46. RAF:	1	sim levam rapazes põem ao lado e mulher também			/
	3	que tem os rapazes e os rapazes também não posso entrar o			//
52. RAF:	1	é bonito por ver – os rapazes – rapazes os os			//
	2	mu//esse que sabe mais/os rapazes também que vão/que pode			/
86. RAF:	1	eu outras pessoas só oiço dizer não recebem bem		/	
96. RAF:	5	que teu pai quando te chamar tem que reponter com respeito		//	
102. RAF:	7	de pessoas é falta de respeito/de respeito		//	
108. RAF:	3	quê mãe que deixou vão lá dentro roubam dinheiro isso já			/
	6	responder muito mal dizer «você não tem nada			/
158. RAF:	6	acabar todas aquelas pedra restar um ali na cova mesmo assim		/	
170. RAF:	3	bonitos você não apanha sujo na rua mas aqui você apanha		/	
180. RAF:	2	pessoas quando vão a um lugar de repente chega muito cedo		/	
			TOTAL	**10**	**8**

APPENDIX 2: CASE TWO – DISTRIBUTION OF TOKENS

CASE TWO

1 = Middle flap [ɾ]

or

0 = Middle trill [r] or uvular [ʁ]

Identification of the participant: MF12RAF

Type of data: Recorded Interview (Digitalized)

Total of tokens: 17

String code: (1SPFy

Middle /ʁ/ = Trill or uvular
"rr" (*erre multiple*)

			T/Tokens	Flap	N/Flap
Speech No.	**No.**	**Line**	**17**	**1**	**0**
16. RAF:	2	horrível não posso namorar – tem que uma pessoa arranjar		//	
30. RAF:	4	o vestido de casamento - podem usar capulanas amarram/já		/	
50. RAF:	3	uma coisa horrível não posso fazer aquilo nem aquilo já--		/	
	5	aqui na... já começa escorregar óleo até nariz sem se m cair			/
70. RAF:	2	tifaqui nabarraça			/
110. RAF:	2	loiça varrer ir na/ir na mes//na escola estudar voltar e		/	
118. RAF:	4	pôr àrvore tinham arroz outro... feijão tudo			/
126. RAF:	1	correram bem nunca não fui atrasar mas agora		/	
130. RAF:	1	não sei não sei porque eu sou burra acho que é			/
136. RAF:	9	coisas muito errada ele começa dizer - - ensinar		/	
156. RAF:	1	arranja pedra pode ser doze cavas começa a jogar		/	
162. RAF:	4	de féria fico em casa a cozinhar: a lavar loiça varrer		/	
	5	arrumar tudo toda coisa que fazem dentro da casa eu faço		/	
178. RAF:	2	porque acontece um dia «vai esses carros que andam por aí		/	
182. RAF	2	carro muito carro e também para toda a gente um lá/para		//	
			TOTAL	**13**	**4**

APPENDIX 3: CASE THREE – DISTRIBUTION OF TOKENS

CASE THREE

1 = Middle trill [r] or uvular [ʁ]

or

0 = Middle flap [ɾ]

Identification of the participant: MR12RAF

Type of data: Recorded Interview (digitalized)

Total of tokens: 100

String code: (1SPFy

Middle /ʁ/ = Flap [ɾ]

			T/Tokens	Trill/Uv	Não Trill/Uv
Speech No.	**No.**	**Line**	**100**	**1**	**0**
6. RAF:	1	ela namoram mas eu não meu pai não me deixa			/
10. RAF:	1	não sei não sei--(interrupção) eu acho namorar é			/
	3	não atendem nada passam namorar dizer «eu vou namorar			//
	4	porque quero tratar minha vida minha mãe não posso se meter			/
12. RAF:	1	elas passem namorar quan//com/o homem ela quer			/
	2	serem a mãe pode falar coisas dizer «eu não quero eu quero			///
	3	homem eu é que amo que gosto do meu coração» já elas se			/
14. RAF:	2	namorar sim			/
16. RAF:	1	namoro? não sei não posso dizer eu acho uma coisa			/
	2	horrível não posso namorar – tem que uma pessoa arranjar			/
	4	possa ajudar quando você o seu futuro			/
22. RAF:	2	passear vem uma hora de madrugada o marido não há-de gostar		/	
	4	e outr//outras pessoas o marido saiem logo às dezassete até			/
	5	voldam quatro hora da manhã enquanto estão grossas isso			/
24. RAF	1	eu gostaria ficar cum//cumprir a ordem do marido			//
	2	ficar em casa tudo o marido que vai dizer ela cumpriu		/	
26. RAF:	1	macua sim é diferente-- aqui para casar tem que			//
30. RAF	1	a cerimónia há só vem -- trazem-- podem preparar			//
	2	uma coisa para comer trazem muitas coisas eles é que sabem			/
	3	que podem servir à mesa já começam a preparar a roupa - não			/
	6	entender no coração dele para os dois já – começam ficam			/
	7	todos a dançar outro vai para casa enfim outras coisa já			/
34. RAF	1	criança quando faz/quando nasce até com quarenta		/	
	2	dia dentro da casa sem sair para fora sem ninguém de fora			///
	3	da da possem tocar quando vai sair para fora fazem isso			//
	5	dançar levam pépé põem na peneira começam a dançar ...			/
	6	já que possam tirar para fora			//
38. RAF:	1	nada diga eu que não podem ensinar agora porque			/
	2	nós não temos idade para ensinar/elas nós não pode ensinar			/
			TOTAL	**5**	**95**

APPENDIX 4: CASE FOUR – DISTRIBUTION OF TOKENS

CASE FOUR

1 = Middle lateral [l] or fricative [ɬ]

or 0 = Middle palatal [ʎ]

Participant's identification: MF12RAF

Type of data: Recorded Interview (Digitalized)

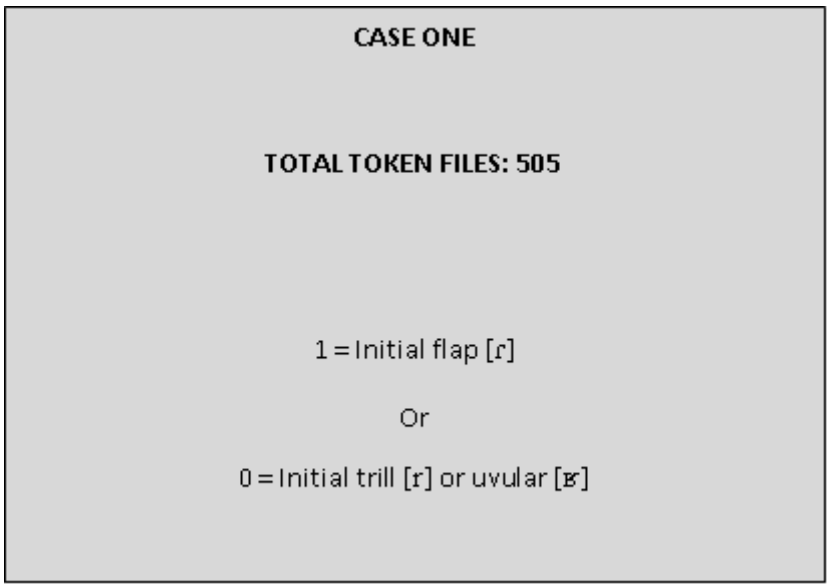
Total of tokens: 25

String code: (1SPFy

Middle /ʎ/ = Lateral palatal [l] or fricative [ɬ]

			T/Tokens	L/Fri	N-L/Fric
Speech No.	**No.**	**Line**	**25**	**1**	**0**
16. RAF:	3	homem que trabalha que tem que tem um bom trabalho ...			//
18. RAF:	2	vinte também eu acho que é melhor		/	
22. RAF:	2	idade começam a discutir outra a mu//a mulher sai viu		/	
36. RAF:	1	este que se entende são mais velhas e mais velhos é		//	
46. RAF:	1	sim levam rapazes põem ao lado e mulher também põem		/	
	2	outra casa posso nem/a mulher não posso entrar o lugar que		/	
	4	que tem mulheres até quarenta dias começam/saiem para ...		/	
50. RAF:	1	a festa das meninas preparam as mais velha começam		/	
52. RAF:	3	saber – não podem assistir só mais velhas é que podem ir		/	
66. RAF:	1	senti-me ma//um pouco melhor porque já outras coisa			/
82. RAF:	2	homens só que trabalha os mulheres também não é proibido		/	/
	3	também tra//trabalhar		/	
100. RAF:	7	não são minhas filhas que não são meus filhos» eu acho que		//	
108. RAF:	7	ver comigo minha mãe que vai trabalhar trazer dinheiro em		/	
	9	trabalha//não me deixava brincar com as crianças - - era por		/	
142. RAF:	2	que vivem como é que trabalham e como é que estudam?		/	
176. RAF:	3	pai casa outra mulher as crianças fica sem entender com			/
186. RAF:	2	sim à espera de nós minha tia meu tio mais os filhos mais		/	
188. RAF:	1	elas trabalham a minha tia é – minha tia – ela		/	
	2	trabalha - - é um piloto não sei pilota que anda por aí no		/	
	3	avião o meu tio trabalha na na fábrica de... já me esqueci o		/	
			TOTAL	**20**	**5**

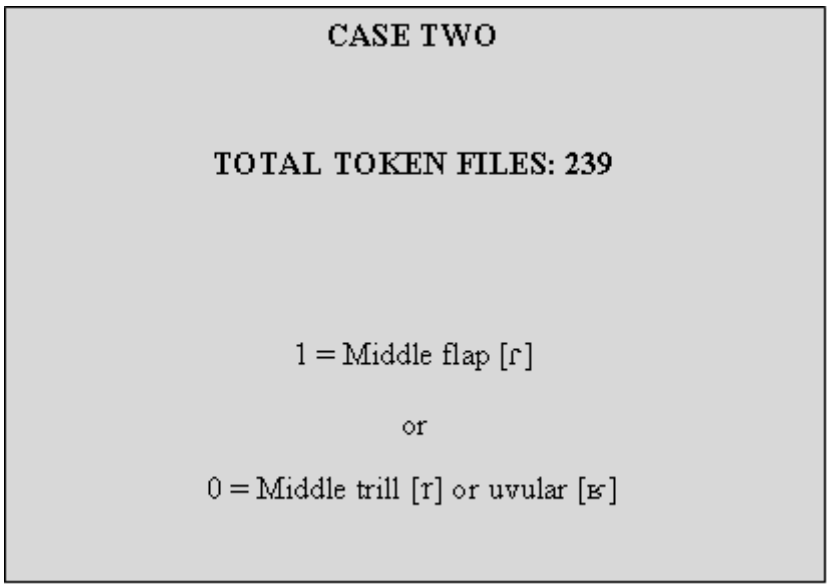
APPENDIX 5: CASE ONE – TOKEN FILES



Initial /ʁ/ = Flap [ɾ]

(1SPFy
(1SPFy
(1SPFy
(1SPFy
(1SPFy
(1SPFy
(1SPFy
(1SPFy
(1SPFy
(1SPFy
(0SPFy
(0SPFy
(0SPFy
(0SPFy
(0SPFy

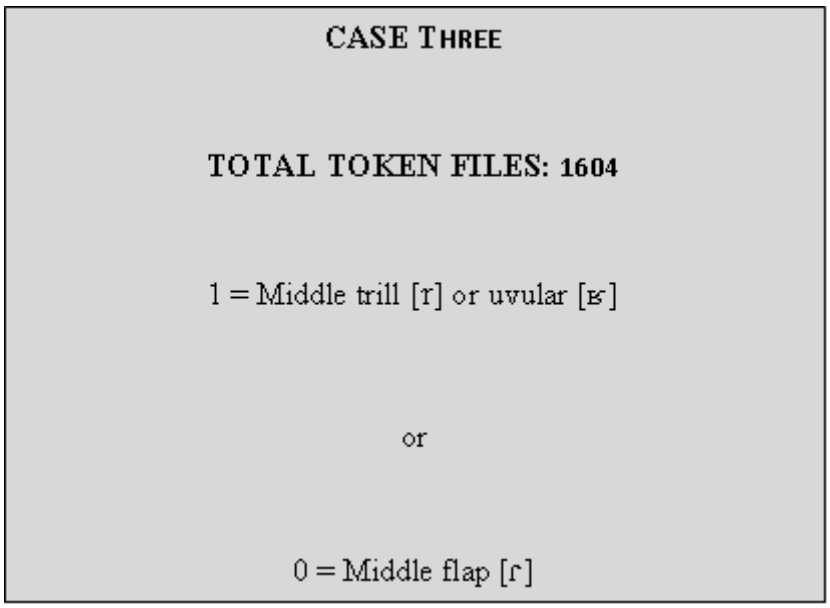
APPENDIX 6: CASE TWO - TOKEN FILES



Middle /ʀ/ (“rr” multiple erre) = Flap [ɾ]

(1SPFy
(1SPFy
(1SPFy
(1SPFy
(1SPFy
(1SPFy
(1SPFy
(1SPFy
(1SPFy
(1SPFy
(1SPFy
(1SPFy
(1SPFy
(1SPFy
(0SPFy
(0SPFy

APPENDIX 7: CASE THREE - TOKEN FILES



Middle /ʀ/ = Middle trill [r] or [ʁ]

(1SPFy

(1SPFy

(1SPFy

(1SPFy

(1SPFy

(1UPFy Dummy

(0SPFy

(0SPFy

(0SPFy

(0SPFy

(0SPFy

(0SPFy

(0SPFy

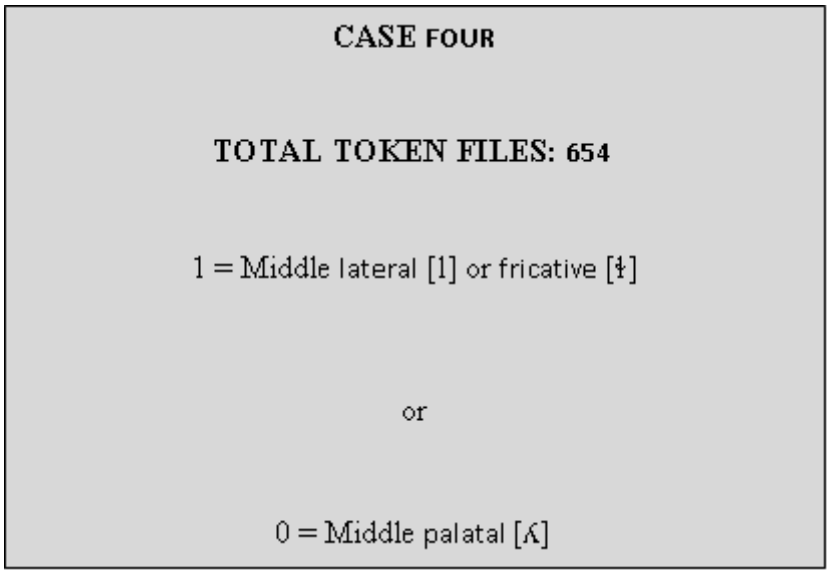
(0SPFy

(0SPFy

(0SPFy

(0SPFy

APPENDIX 8: CASE FOUR – TOKEN FILES



Middle /ʎ/ = Lateral laterl [l] or fricative [ɬ]

(1SPFy

(1SPFy

(1SPFy

(1SPFy

(1SPFy

(1SPFy

(1SPFy

(1SPFy

(1SPFy

(1SPFy

(1SPFy

(1SPFy

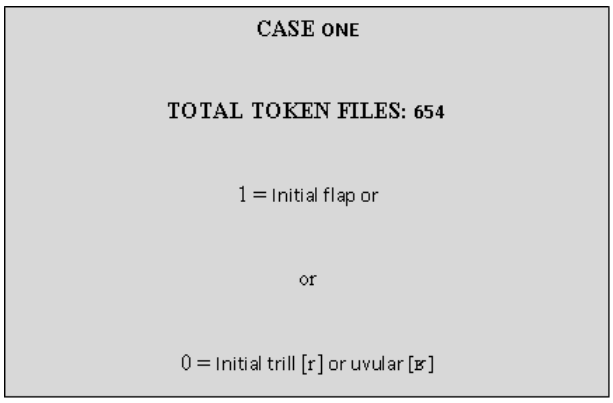
(1SPFy

(1SPFy

(1SPFy

(1SPFy

APPENDIX 9: CASE ONE – GOLDVARB FIRST RUN

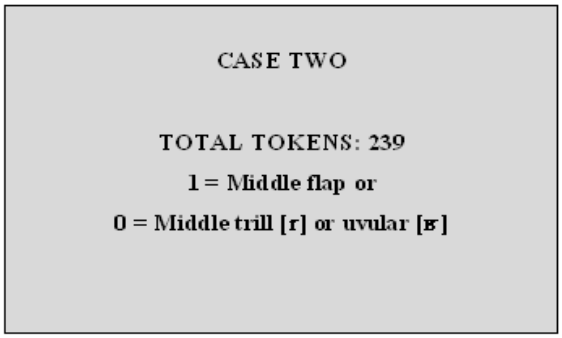


CELL CREATION

```
=============
    Name of token file: Untitled.tkn
Name of condition file: Untitled.cnd
(
(1)
(2)
(3)
(4)
(5)
)
       Number of cells:  10
  Application value(s):  10
  Total no. of factors:  9

 Group        1      0   Total  %
---------------------------------
 1 (2)
   S   N    265    132     397  78
       %     66     33

   U   N     94     14     108  21
       %     87     12

 Total N    359    146     505
       %     71     28
---------------------------------
```

```
 2 (3)
   P   N    222     97     319  63
       %     69     30

   H   N    137     49     186  36
       %     73     26

 Total N    359    146     505
       %     71     28
---------------------------------
 3 (4)
   F   N    186     74     260  51
       %     71     28

   M   N    173     72     245  48
       %     70     29

 Total N    359    146     505
       %     71     28
---------------------------------
 4 (5)
   y   N    180     99     279  55
       %     64     35

   m   N     87     40     127  25
       %     68     31

   o   N     92      7      99  19
       %     92      7

 Total N    359    146     505
       %     71     28
---------------------------------
 Total N    359    146     505
       %     71     28

 Name of new cell file: Untitled.cel
```

APPENDIX 10: CASE TWO – GOLDVARB FIRST RUN



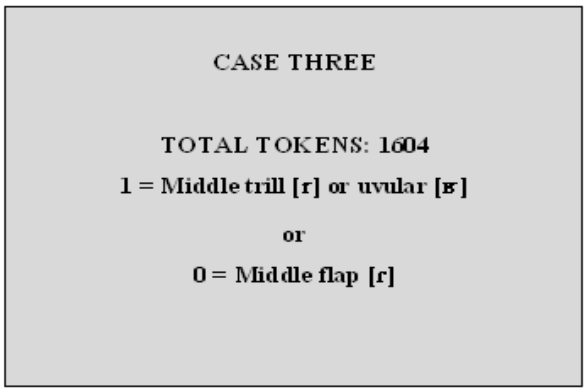
CELL CREATION

```
=============
    Name of token file: Untitled.tkn
Name of condition file: Untitled.cnd
(
(1)
(2)
(3)
(4)
(5)
)
       Number of cells:  9
  Application value(s):  10
  Total no. of factors:  9

 Group        1      0   Total  %
----------------------------------
 1 (2)
   S   N    179     45     224  93
       %     79     20

   U   N     14      1      15   6
       %     93      6

 Total N    193     46     239
       %     80     19
----------------------------------
```

```
2 (3)
  P   N    143     35     178  74
      %     80     19

  H   N     50     11      61  25
      %     81     18

Total N    193     46     239
      %     80     19
---------------------------------
3 (4)
  F   N     85     24     109  45
      %     77     22

  M   N    108     22     130  54
      %     83     16

Total N    193     46     239
      %     80     19
---------------------------------
4 (5)
  y   N    128     37     165  69
      %     77     22

  m   N     18      5      23   9
      %     78     21

  o   N     47      4      51  21
      %     92      7

Total N    193     46     239
      %     80     19
---------------------------------
Total N    193     46     239
      %     80     19
```

Name of new cell file: Untitled.cel

APPENDIX 11: CASE THREE – GOLDVARB FIRST RUN

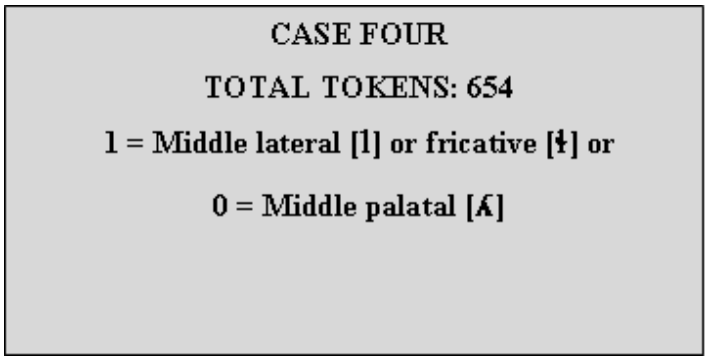


CELL CREATION

```
=============
    Name of token file: Untitled.tkn
Name of condition file: Untitled.cnd
(
(1)
(2)
(3)
(4)
(5)
)
       Number of cells:  12
  Application value(s):  10
  Total no. of factors:  9

 Group        1      0   Total  %
---------------------------------
 1 (2)
   S   N     18   1285    1303  81
       %      1     98

   U   N      2    301     303  18
       %      0     99

 Total N     20   1586    1606
       %      1     98
---------------------------------
```

```
 2 (3)
   P   N     13    990    1003  62
       %      1     98

   H   N      7    596     603  37
       %      1     98

 Total N     20   1586    1606
       %      1     98
---------------------------------
 3 (4)
   F   N      8    695     703  43
       %      1     98

   M   N     12    891     903  56
       %      1     98

 Total N     20   1586    1606
       %      1     98
---------------------------------
 4 (5)
   y   N     17    886     903  56
       %      1     98

   m   N      1    401     402  25
       %      0     99

   o   N      2    299     301  18
       %      0     99

 Total N     20   1586    1606
       %      1     98
---------------------------------
 Total N     20   1586    1606
       %      1     98

 Name of new cell file:Untitled.cel
```

APPENDIX 12: CASE FOUR – GOLDVARB FIRST RUN

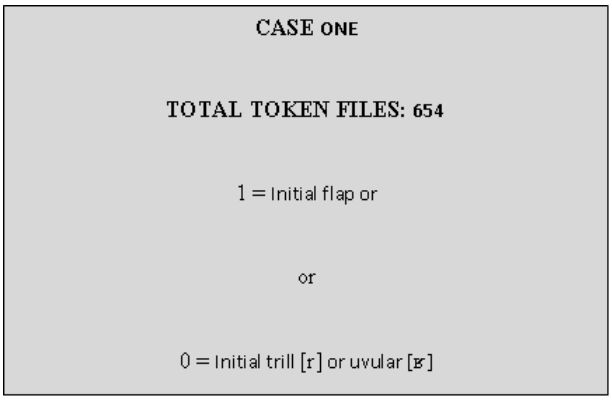


CELL CREATION

```
=============
    Name of token file: Untitled.tkn
Name of condition file: Untitled.cnd
(
(1)
(2)
(3)
(4)
(5)
)
       Number of cells:  9
  Application value(s):  10
  Total no. of factors:  9

 Group        1      0   Total  %
---------------------------------
 1 (2)
   S   N    397    179     576  93
       %     68     31

   U   N     32      6      38   6
       %     84     15

 Total N    429    185     614
       %     69     30
---------------------------------
```

```
 2 (3)
   P   N    315    142     457  74
       %     68     31

   H   N    114     43     157  25
       %     72     27

 Total N    429    185     614
       %     69     30
---------------------------------
 3 (4)
   F   N    232     62     294  47
       %     78     21

   M   N    197    123     320  52
       %     61     38

 Total N    429    185     614
       %     69     30
---------------------------------
 4 (5)
   y   N    235    139     374  60
       %     62     37

   m   N    110     15     125  20
       %     88     12

   o   N     84     31     115  18
       %     73     26

 Total N    429    185     614
       %     69     30
---------------------------------
 Total N    429    185     614
       %     69     30

 Name of new cell file: Untitled.cel
```

APPENDIX 13: CASE ONE – BINOMIAL VARBRUL



Binomial VarbRul

```
================
Name of cell file: Untitled.cel

Using fast, less accurate method.
Averaging by weighting factors.
Threshold, step-up/down: 0.050001

 # Stepping up:
 # Stepping up:

---------- Level # 0 ----------

Run # 1, 1 cells:
Convergence at Iteration 2
Input 0.711
Log likelihood = -303.683

---------- Level # 1 ----------

Run # 2, 2 cells:
Convergence at Iteration 5
Input 0.722
Group # 1 -- S: 0.436, U: 0.721
Log likelihood = -294.118 Significance = 0.000


Run # 3, 2 cells:
Convergence at Iteration 4
Input 0.711
```

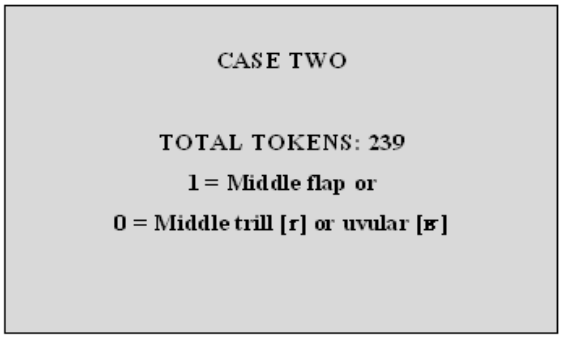
```
Group # 2 -- P: 0.482, H: 0.531
Log likelihood = -303.207 Significance = 0.340

Run # 4, 2 cells:
Convergence at Iteration 3
Input 0.711
Group # 3 -- F: 0.505, M: 0.495
Log likelihood = -303.656 Significance = 0.824

Run # 5, 3 cells:
Convergence at Iteration 5
Input 0.737
Group # 4 -- y: 0.394, m: 0.437, o: 0.824
Log likelihood = -285.873 Significance = 0.000

Add Group # 4 with factors ymo

---------- Level # 2 ----------

Run # 6, 6 cells:
Convergence at Iteration 6
Input 0.742
Group # 1 -- S: 0.457, U: 0.654
Group # 4 -- y: 0.419, m: 0.418, o: 0.793
Log likelihood = -282.673 Significance = 0.012

Run # 7, 6 cells:
Convergence at Iteration 5
Input 0.737
Group # 2 -- P: 0.491, H: 0.515
Group # 4 -- y: 0.392, m: 0.443, o: 0.822
Log likelihood = -285.786 Significance = 0.684

Run # 8, 6 cells:
Convergence at Iteration 6
Input 0.738
Group # 3 -- F: 0.452, M: 0.551
Group # 4 -- y: 0.380, m: 0.442, o: 0.844
Log likelihood = -284.059 Significance = 0.059

Add Group # 1 with factors SU

---------- Level # 3 ----------

Run # 9, 8 cells:
Convergence at Iteration 6
```

```
Input 0.742
Group # 1 -- S: 0.457, U: 0.654
Group # 2 -- P: 0.501, H: 0.498
Group # 4 -- y: 0.420, m: 0.417, o: 0.793
Log likelihood = -282.673 Significance = 0.977

Run # 10, 9 cells:
Convergence at Iteration 9
Input 0.741
Group # 1 -- S: 0.463, U: 0.632
Group # 3 -- F: 0.471, M: 0.530
Group # 4 -- y: 0.407, m: 0.425, o: 0.809
Log likelihood = -282.101 Significance = 0.288

No remaining groups significant

Groups selected while stepping up:  4  1
Best stepping up run:   #6
---------------------------------------------

 # Stepping down:
 # Stepping down:

---------- Level # 4 ----------

Run # 11, 10 cells:
Convergence at Iteration 10
Input 0.741
Group # 1 -- S: 0.464, U: 0.631
Group # 2 -- P: 0.518, H: 0.469
Group # 3 -- F: 0.459, M: 0.544
Group # 4 -- y: 0.406, m: 0.418, o: 0.816
Log likelihood = -281.823

---------- Level # 3 ----------

Run # 12, 9 cells:
Convergence at Iteration 10
Input 0.739
Group # 2 -- P: 0.518, H: 0.468
Group # 3 -- F: 0.439, M: 0.565
Group # 4 -- y: 0.378, m: 0.431, o: 0.854
Log likelihood = -283.759 Significance = 0.049

Run # 13, 9 cells:
Convergence at Iteration 9
```

```
Input 0.741
Group # 1 -- S: 0.463, U: 0.632
Group # 3 -- F: 0.471, M: 0.530
Group # 4 -- y: 0.407, m: 0.425, o: 0.809
Log likelihood = -282.101 Significance = 0.467

Run # 14, 8 cells:
Convergence at Iteration 6
Input 0.742
Group # 1 -- S: 0.457, U: 0.654
Group # 2 -- P: 0.501, H: 0.498
Group # 4 -- y: 0.420, m: 0.417, o: 0.793
Log likelihood = -282.673 Significance = 0.195

Run # 15, 6 cells:
Convergence at Iteration 5
Input 0.722
Group # 1 -- S: 0.436, U: 0.721
Group # 2 -- P: 0.500, H: 0.500
Group # 3 -- F: 0.510, M: 0.489
Log likelihood = -294.030 Significance = 0.000

Cut Group # 2 with factors PH

---------- Level # 2 ----------



Run # 16, 6 cells:
Convergence at Iteration 6
Input 0.738
Group # 3 -- F: 0.452, M: 0.551
Group # 4 -- y: 0.380, m: 0.442, o: 0.844
Log likelihood = -284.059 Significance = 0.049

Run # 17, 6 cells:
Convergence at Iteration 6
Input 0.742
Group # 1 -- S: 0.457, U: 0.654
Group # 4 -- y: 0.419, m: 0.418, o: 0.793
Log likelihood = -282.673 Significance = 0.288

Run # 18, 4 cells:
Convergence at Iteration 5
Input 0.722
Group # 1 -- S: 0.436, U: 0.721
```

```
Group # 3 -- F: 0.510, M: 0.489
Log likelihood = -294.030 Significance = 0.000

Cut Group # 3 with factors FM

---------- Level # 1 ----------

Run # 19, 3 cells:
Convergence at Iteration 5
Input 0.737
Group # 4 -- y: 0.394, m: 0.437, o: 0.824
Log likelihood = -285.873 Significance = 0.012

Run # 20, 2 cells:
Convergence at Iteration 5
Input 0.722
Group # 1 -- S: 0.436, U: 0.721
Log likelihood = -294.118 Significance = 0.000

All remaining groups significant

Groups eliminated while stepping down:  2  3
Best stepping  up  run: #6
Best stepping down run: #17
```

APPENDIX 14: CASE TWO – BINOMIAL VARBRUL



Binomial VarbRul

```
================
Name of cell file: Untitled.cel

Using fast, less accurate method.
Averaging by weighting factors.
Threshold, step-up/down: 0.050001

 # Stepping up:
 # Stepping up:

---------- Level # 0 ----------

Run # 1, 1 cells:
Convergence at Iteration 2
Input 0.808
Log likelihood = -117.058

---------- Level # 1 ----------

Run # 2, 2 cells:
Convergence at Iteration 5
Input 0.811
Group # 1 -- S: 0.480, U: 0.765
Log likelihood = -116.041 Significance = 0.163



Run # 3, 2 cells:
Convergence at Iteration 3
Input 0.808
```

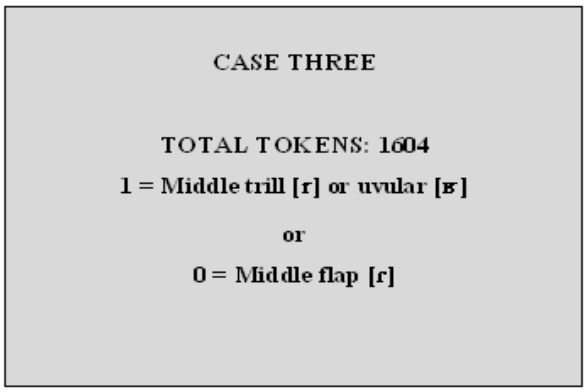
```
Group # 2 -- P: 0.493, H: 0.519
Log likelihood = -117.019 Significance = 0.783

Run # 4, 2 cells:
Convergence at Iteration 4
Input 0.809
Group # 3 -- F: 0.456, M: 0.537
Log likelihood = -116.565 Significance = 0.328

Run # 5, 3 cells:
Convergence at Iteration 5
Input 0.818
Group # 4 -- y: 0.434, m: 0.444, o: 0.722
Log likelihood = -113.881 Significance = 0.044

Add Group # 4 with factors ymo

---------- Level # 2 ----------

Run # 6, 5 cells:
Convergence at Iteration 6
Input 0.820
Group # 1 -- S: 0.487, U: 0.679
Group # 4 -- y: 0.439, m: 0.454, o: 0.706
Log likelihood = -113.550 Significance = 0.433

Run # 7, 6 cells:
Convergence at Iteration 5
Input 0.819
Group # 2 -- P: 0.489, H: 0.531
Group # 4 -- y: 0.434, m: 0.442, o: 0.724
Log likelihood = -113.781 Significance = 0.667

Run # 8, 6 cells:
Convergence at Iteration 5
Input 0.819
Group # 3 -- F: 0.464, M: 0.530
Group # 4 -- y: 0.434, m: 0.460, o: 0.717
Log likelihood = -113.583 Significance = 0.454

No remaining groups significant

Groups selected while stepping up:  4
Best stepping up run:   #5
-------------------------------------------
```

```
 # Stepping down:
 # Stepping down:

---------- Level # 4 ----------

Run # 9, 9 cells:
Convergence at Iteration 5
Input 0.821
Group # 1 -- S: 0.485, U: 0.713
Group # 2 -- P: 0.511, H: 0.469
Group # 3 -- F: 0.453, M: 0.539
Group # 4 -- y: 0.442, m: 0.479, o: 0.688
Log likelihood = -113.179

---------- Level # 3 ----------


Run # 10, 8 cells:
Convergence at Iteration 5
Input 0.819
Group # 2 -- P: 0.498, H: 0.506
Group # 3 -- F: 0.466, M: 0.528
Group # 4 -- y: 0.434, m: 0.458, o: 0.718
Log likelihood = -113.579 Significance = 0.389

Run # 11, 8 cells:
Convergence at Iteration 6
Input 0.821
Group # 1 -- S: 0.487, U: 0.683
Group # 3 -- F: 0.463, M: 0.531
Group # 4 -- y: 0.440, m: 0.471, o: 0.697
Log likelihood = -113.236 Significance = 0.743

Run # 12, 7 cells:
Convergence at Iteration 7
Input 0.820
Group # 1 -- S: 0.488, U: 0.668
Group # 2 -- P: 0.496, H: 0.512
Group # 4 -- y: 0.438, m: 0.452, o: 0.708
Log likelihood = -113.537 Significance = 0.416

Run # 13, 5 cells:
Convergence at Iteration 10
Input 0.814
Group # 1 -- S: 0.473, U: 0.838
Group # 2 -- P: 0.529, H: 0.416
```

```
Group # 3 -- F: 0.425, M: 0.563
Log likelihood = -114.925 Significance = 0.180

Cut Group # 2 with factors PH

---------- Level # 2 ----------

Run # 14, 6 cells:
Convergence at Iteration 5
Input 0.819
Group # 3 -- F: 0.464, M: 0.530
Group # 4 -- y: 0.434, m: 0.460, o: 0.717
Log likelihood = -113.583 Significance = 0.423

Run # 15, 5 cells:
Convergence at Iteration 6
Input 0.820
Group # 1 -- S: 0.487, U: 0.679
Group # 4 -- y: 0.439, m: 0.454, o: 0.706
Log likelihood = -113.550 Significance = 0.444

Run # 16, 4 cells:
Convergence at Iteration 6
Input 0.813
Group # 1 -- S: 0.479, U: 0.779
Group # 3 -- F: 0.449, M: 0.543
Log likelihood = -115.390 Significance = 0.121

Cut Group # 3 with factors FM

---------- Level # 1 ----------

Run # 17, 3 cells:
Convergence at Iteration 5
Input 0.818
Group # 4 -- y: 0.434, m: 0.444, o: 0.722
Log likelihood = -113.881 Significance = 0.433

Run # 18, 2 cells:
Convergence at Iteration 5
Input 0.811
Group # 1 -- S: 0.480, U: 0.765
Log likelihood = -116.041 Significance = 0.086

Cut Group # 1 with factors SU
```

```
---------- Level # 0 ----------

Run # 19, 1 cells:
Convergence at Iteration 2
Input 0.808
Log likelihood = -117.058 Significance = 0.044

All remaining groups significant

Groups eliminated while stepping down:  2  3  1
Best stepping  up  run: #5
Best stepping down run: #17
```

APPENDIX 15: CASE THREE – BINOMIAL VARBRUL



Binomial VarbRul

```
================
Name of cell file: Untitled.cel

Using fast, less accurate method.
Averaging by weighting factors.
Threshold, step-up/down: 0.050001

 # Stepping up:
 # Stepping up:

---------- Level # 0 ----------

Run # 1, 1 cells:
Convergence at Iteration 2
Input 0.012
Log likelihood = -107.590

---------- Level # 1 ----------

Run # 2, 2 cells:
Convergence at Iteration 5
Input 0.012
Group # 1 -- S: 0.535, U: 0.354
Log likelihood = -106.987 Significance = 0.277

Run # 3, 2 cells:
Convergence at Iteration 3
Input 0.012
Group # 2 -- P: 0.510, H: 0.483
Log likelihood = -107.562 Significance = 0.816
```

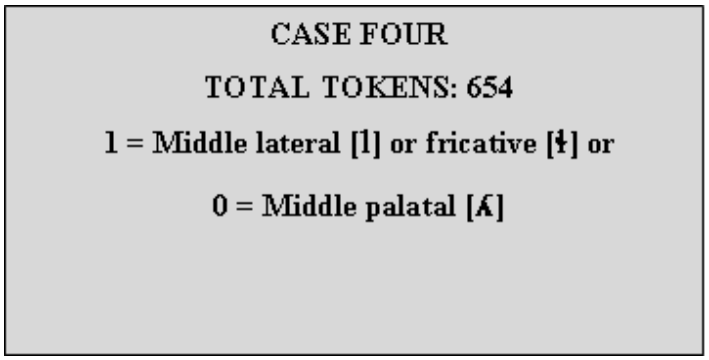
```
Run # 4, 2 cells:
Convergence at Iteration 3
Input 0.012
Group # 3 -- F: 0.478, M: 0.517
Log likelihood = -107.531 Significance = 0.736

Run # 5, 3 cells:
Convergence at Iteration 5
Input 0.009
Group # 4 -- y: 0.670, m: 0.209, o: 0.414
Log likelihood = -103.388 Significance = 0.016

Add Group # 4 with factors ymo

---------- Level # 2 ----------

Run # 6, 6 cells:
Convergence at Iteration 6
Input 0.009
Group # 1 -- S: 0.520, U: 0.416
Group # 4 -- y: 0.664, m: 0.213, o: 0.428
Log likelihood = -103.222 Significance = 0.580

Run # 7, 6 cells:
Convergence at Iteration 6
Input 0.009
Group # 2 -- P: 0.530, H: 0.450
Group # 4 -- y: 0.675, m: 0.203, o: 0.411
Log likelihood = -103.147 Significance = 0.491

Run # 8, 6 cells:
Convergence at Iteration 6
Input 0.009
Group # 3 -- F: 0.515, M: 0.488
Group # 4 -- y: 0.672, m: 0.208, o: 0.409
Log likelihood = -103.360 Significance = 0.815

No remaining groups significant

Groups selected while stepping up:  4
Best stepping up run:   #5
-----------------------------------------------

 # Stepping down:
 # Stepping down:
```

```
---------- Level # 4 ----------

Run # 9, 12 cells:
Convergence at Iteration 7
Input 0.009
Group # 1 -- S: 0.512, U: 0.448
Group # 2 -- P: 0.527, H: 0.455
Group # 3 -- F: 0.491, M: 0.507
Group # 4 -- y: 0.669, m: 0.208, o: 0.420
Log likelihood = -103.090

---------- Level # 3 ----------

Run # 10, 9 cells:
Convergence at Iteration 6
Input 0.009
Group # 2 -- P: 0.534, H: 0.443
Group # 3 -- F: 0.489, M: 0.508
Group # 4 -- y: 0.674, m: 0.203, o: 0.414
Log likelihood = -103.137 Significance = 0.764

Run # 11, 11 cells:
Convergence at Iteration 6
Input 0.009
Group # 1 -- S: 0.519, U: 0.418
Group # 3 -- F: 0.510, M: 0.492
Group # 4 -- y: 0.665, m: 0.212, o: 0.424
Log likelihood = -103.210 Significance = 0.640

Run # 12, 9 cells:
Convergence at Iteration 8
Input 0.009
Group # 1 -- S: 0.512, U: 0.447
Group # 2 -- P: 0.524, H: 0.461
Group # 4 -- y: 0.670, m: 0.208, o: 0.418
Log likelihood = -103.097 Significance = 0.908

Run # 13, 7 cells:
Convergence at Iteration 6
Input 0.012
Group # 1 -- S: 0.535, U: 0.353
Group # 2 -- P: 0.509, H: 0.485
Group # 3 -- F: 0.465, M: 0.527
Log likelihood = -106.870 Significance = 0.024
```

```
Cut Group # 3 with factors FM

---------- Level # 2 ----------

Run # 14, 6 cells:
Convergence at Iteration 6
Input 0.009
Group # 2 -- P: 0.530, H: 0.450
Group # 4 -- y: 0.675, m: 0.203, o: 0.411
Log likelihood = -103.147 Significance = 0.758

Run # 15, 6 cells:
Convergence at Iteration 6
Input 0.009
Group # 1 -- S: 0.520, U: 0.416
Group # 4 -- y: 0.664, m: 0.213, o: 0.428
Log likelihood = -103.222 Significance = 0.634

Run # 16, 4 cells:
Convergence at Iteration 5
Input 0.012
Group # 1 -- S: 0.535, U: 0.352
Group # 2 -- P: 0.498, H: 0.504
Log likelihood = -106.985 Significance = 0.021

Cut Group # 1 with factors SU

---------- Level # 1 ----------

Run # 17, 3 cells:
Convergence at Iteration 5
Input 0.009
Group # 4 -- y: 0.670, m: 0.209, o: 0.414
Log likelihood = -103.388 Significance = 0.491

Run # 18, 2 cells:
Convergence at Iteration 3
Input 0.012
Group # 2 -- P: 0.510, H: 0.483
Log likelihood = -107.562 Significance = 0.013

Cut Group # 2 with factors PH

---------- Level # 0 ----------

Run # 19, 1 cells:
```

```
Convergence at Iteration 2
Input 0.012
Log likelihood = -107.590 Significance = 0.016

All remaining groups significant

Groups eliminated while stepping down:  3  1  2
Best stepping  up  run: #5
Best stepping down run: #17
```

APPENDIX 16: CASE FOUR – BINOMIAL VarbRul



Binomial VarbRul

```
================
Name of cell file: Untitled.cel

Using fast, less accurate method.
Averaging by weighting factors.
Threshold, step-up/down: 0.050001

 # Stepping up:
 # Stepping up:

---------- Level # 0 ----------

Run # 1, 1 cells:
Convergence at Iteration 2
Input 0.699
Log likelihood = -375.746

---------- Level # 1 ----------

Run # 2, 2 cells:
Convergence at Iteration 5
Input 0.701
Group # 1 -- S: 0.486, U: 0.695
Log likelihood = -373.527 Significance = 0.039

Run # 3, 2 cells:
Convergence at Iteration 4
Input 0.699
Group # 2 -- P: 0.489, H: 0.533
Log likelihood = -375.365 Significance = 0.401
```

```
Run # 4, 2 cells:
Convergence at Iteration 4
Input 0.706
Group # 3 -- F: 0.608, M: 0.400
Log likelihood = -364.620 Significance = 0.000

Run # 5, 3 cells:
Convergence at Iteration 5
Input 0.713
Group # 4 -- y: 0.405, m: 0.746, o: 0.521
Log likelihood = -359.668 Significance = 0.000

Add Group # 4 with factors ymo

---------- Level # 2 ----------

Run # 6, 5 cells:
Convergence at Iteration 6
Input 0.716
Group # 1 -- S: 0.485, U: 0.713
Group # 4 -- y: 0.410, m: 0.756, o: 0.491
Log likelihood = -357.142 Significance = 0.027

Run # 7, 6 cells:
Convergence at Iteration 5
Input 0.714
Group # 2 -- P: 0.484, H: 0.548
Group # 4 -- y: 0.403, m: 0.749, o: 0.524
Log likelihood = -358.897 Significance = 0.217

Run # 8, 6 cells:
Convergence at Iteration 6
Input 0.716
Group # 3 -- F: 0.574, M: 0.432
Group # 4 -- y: 0.428, m: 0.714, o: 0.487
Log likelihood = -355.687 Significance = 0.007

Add Group # 3 with factors FM
---------- Level # 3 ----------

Run # 9, 8 cells:
Convergence at Iteration 6
Input 0.718
Group # 1 -- S: 0.486, U: 0.705
Group # 3 -- F: 0.571, M: 0.434
Group # 4 -- y: 0.432, m: 0.724, o: 0.461
```

```
Log likelihood = -353.394 Significance = 0.036

Run # 10, 8 cells:
Convergence at Iteration 6
Input 0.717
Group # 2 -- P: 0.467, H: 0.594
Group # 3 -- F: 0.595, M: 0.413
Group # 4 -- y: 0.432, m: 0.703, o: 0.488
Log likelihood = -353.009 Significance = 0.021

Add Group # 2 with factors PH

---------- Level # 4 ----------

Run # 11, 9 cells:
Convergence at Iteration 7
Input 0.718
Group # 1 -- S: 0.491, U: 0.637
Group # 2 -- P: 0.476, H: 0.568
Group # 3 -- F: 0.588, M: 0.419
Group # 4 -- y: 0.433, m: 0.712, o: 0.473
Log likelihood = -352.311 Significance = 0.243

No remaining groups significant

Groups selected while stepping up:  4  3  2
Best stepping up run:   #10
-------------------------------------------

 # Stepping down:
 # Stepping down:

---------- Level # 4 ----------
Run # 12, 9 cells:
Convergence at Iteration 7
Input 0.718
Group # 1 -- S: 0.491, U: 0.637
Group # 2 -- P: 0.476, H: 0.568
Group # 3 -- F: 0.588, M: 0.419
Group # 4 -- y: 0.433, m: 0.712, o: 0.473
Log likelihood = -352.311

---------- Level # 3 ----------

Run # 13, 8 cells:
Convergence at Iteration 6
```

```
Input 0.717
Group # 2 -- P: 0.467, H: 0.594
Group # 3 -- F: 0.595, M: 0.413
Group # 4 -- y: 0.432, m: 0.703, o: 0.488
Log likelihood = -353.009 Significance = 0.243

Run # 14, 8 cells:
Convergence at Iteration 6
Input 0.718
Group # 1 -- S: 0.486, U: 0.705
Group # 3 -- F: 0.571, M: 0.434
Group # 4 -- y: 0.432, m: 0.724, o: 0.461
Log likelihood = -353.394 Significance = 0.151

Run # 15, 7 cells:
Convergence at Iteration 5
Input 0.716
Group # 1 -- S: 0.486, U: 0.703
Group # 2 -- P: 0.496, H: 0.510
Group # 4 -- y: 0.409, m: 0.755, o: 0.493
Log likelihood = -357.113 Significance = 0.004

Run # 16, 5 cells:
Convergence at Iteration 9
Input 0.709
Group # 1 -- S: 0.495, U: 0.569
Group # 2 -- P: 0.465, H: 0.601
Group # 3 -- F: 0.630, M: 0.381
Log likelihood = -360.524 Significance = 0.000

Cut Group # 1 with factors SU

---------- Level # 2 ----------

Run # 17, 6 cells:
Convergence at Iteration 6
Input 0.716
Group # 3 -- F: 0.574, M: 0.432
Group # 4 -- y: 0.428, m: 0.714, o: 0.487
Log likelihood = -355.687 Significance = 0.021

Run # 18, 6 cells:
Convergence at Iteration 5
Input 0.714
Group # 2 -- P: 0.484, H: 0.548
Group # 4 -- y: 0.403, m: 0.749, o: 0.524
```

```
Log likelihood = -358.897 Significance = 0.001

Run # 19, 4 cells:
Convergence at Iteration 7
Input 0.708
Group # 2 -- P: 0.461, H: 0.612
Group # 3 -- F: 0.633, M: 0.378
Log likelihood = -360.692 Significance = 0.000

All remaining groups significant

Groups eliminated while stepping down:  1
Best stepping  up  run: #10
Best stepping down run: #13
```